

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION  
International Bureau



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

<b>(51) International Patent Classification <sup>6</sup> :</b> <b>C12Q 1/68, A01N 43/04, A61K 31/70</b>		<b>A1</b>	<b>(11) International Publication Number:</b> <b>WO 97/37044</b>
			<b>(43) International Publication Date:</b> 9 October 1997 (09.10.97)
<b>(21) International Application Number:</b> PCT/US97/05223			<b>(71) Applicant (for all designated States except US):</b> ASTRA AKTIEBOLAG [SE/SE]; S-151 85 Södertälje (SE).
<b>(22) International Filing Date:</b> 27 March 1997 (27.03.97)			<b>(72) Inventors; and</b> <b>(75) Inventors/Applicants (for US only):</b> SMITH, Douglas [US/US]; 2 Mayflower Lane, Gloucester, MA 01930 (US). ALM, Richard, A. [AU/US]; 28 Russett Hill Road, Ashland, MA 01721 (US).
<b>(30) Priority Data:</b> 08/625,811 29 March 1996 (29.03.96) US 08/758,731 2 April 1996 (02.04.96) US 08/736,905 25 October 1996 (25.10.96) US 08/738,859 28 October 1996 (28.10.96) US 08/761,318 6 December 1996 (06.12.96) US			<b>(74) Agents:</b> MANDRAGOURAS, Amy, E. et al.; Lahive & Cockfield, L.L.P., 28 State Street, Boston, MA 02109 (US).
<b>(60) Parent Applications or Grants</b> <b>(63) Related by Continuation</b> US 08/625,811 (CIP) Filed on 29 March 1996 (29.03.96) US 08/738,859 (CIP) Filed on 28 October 1996 (28.10.96) US 08/761,318 (CIP) Filed on 6 December 1996 (06.12.96) US 08/758,731 (CIP) Filed on 2 April 1996 (02.04.96) US 08/736,905 (CIP) Filed on 25 October 1996 (25.10.96)			<b>(81) Designated States:</b> AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CZ, DE, DK, EE, ES, FI, GB, GE, GH, HU, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ARIPO patent (GH, KE, LS, MW, SD, SZ, UG), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).
			<b>Published</b> <i>With international search report. Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i>
<b>(54) Title:</b> NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO HELICOBACTER PYLORI AND VACCINE COMPOSITIONS THEREOF			
<b>(57) Abstract</b>  Recombinant or substantially pure preparations of <i>H. pylori</i> polypeptides are described. The nucleic acids encoding the polypeptides also are described. The <i>H. pylori</i> polypeptides are useful in diagnostic and vaccine compositions.			

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Larvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
RJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

## NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO *HELICOBACTER PYLORI* AND VACCINE COMPOSITIONS THEREOF

### Background of the Invention

5       *Helicobacter pylori* is a gram-negative, S-shaped, microaerophilic bacterium that was discovered and cultured from a human gastric biopsy specimen. (Warren, J.R. and B. Marshall, (1983) *Lancet* 1: 1273-1275; and Marshall et al., (1984) *Microbios Lett.* 25: 83-88). *H. pylori* has been strongly linked to chronic gastritis and duodenal ulcer disease. (Rathbone et. al., (1986) *Gut* 27: 635-641). Moreover, evidence is  
10 accumulating for an etiologic role of *H. pylori* in nonulcer dyspepsia, gastric ulcer disease, and gastric adenocarcinoma. (Blaser M. J., (1993) *Trends Microbiol.* 1: 255-260). Transmission of the bacteria occurs via the oral route, and the risk of infection increases with age. (Taylor, D.N. and M. J. Blaser, (1991) *Epidemiol. Rev* 13: 42-50). *H. pylori* colonizes the human gastric mucosa, establishing an infection that usually  
15 persists for decades. Infection by *H. pylori* is prevalent worldwide. Developed countries have infection rates over 50% of the adult population, while developing countries have infection rates reaching 90% of the adults over the age of 20. (Hopkins R. J. and J. G. Morris (1994) *Am. J. Med.* 97: 265-277).

      The bacterial factors necessary for colonization of the gastric environment, and  
20 for virulence of this pathogen, are poorly understood. Examples of the putative virulence factors include the following: urease, an enzyme that may play a role in neutralizing gastric acid pH (Eaton et al., (1991) *Infect. Immunol.* 59: 2470-2475; Ferrero, R.L. and A. Lee (1991) *Microb. Ecol. Hlth. Dis.* 4: 121-134; Labigne et al., (1991) *J. Bacteriol.* 173: 1920-1931); the bacterial flagellar proteins responsible for  
25 motility across the mucous layer. (Hazell et al., (1986) *J. Inf. Dis.* 153: 658-663; Lying et al., (1992) *Mol. Microbiol.* 6: 2863-2874; and Haas et al., (1993) *Mol. Microbiol.* 8: 753-760); Vac A, a bacterial toxin that induces the formation of intracellular vacuoles in epithelial cells (Schmitt, W. and R. Haas, (1994) *Molecular Microbiol.* 12(2): 307-319); and several gastric tissue-specific adhesins. (Boren et al., (1993) *Science* 262: 1892-  
30 1895; Evans et al., (1993) *J. Bacteriol.* 175: 674-683; and Falk et al., (1993) *Proc. Natl. Acad. Sci. USA* 90: 2035-203).

      Numerous therapeutic agents are currently available that eradicate *H. pylori* infections *in vitro*. (Huesca et. al., (1993) *Zbl. Bakt.* 280: 244-252; Hopkins, R. J. and J. G. Morris, supra). However, many of these treatments are suboptimally effective *in vivo*  
35 because of bacterial resistance, altered drug distribution, patient non-compliance or poor drug availability. (Hopkins, R. J. and J. G. Morris, supra). Treatment with antibiotics combined with bismuth are part of the standard regime used to treat *H. pylori* infection.

(Malfertheiner, P. and J. E. Dominguez-Munoz (1993) *Clinical Therapeutics* 15 Supp. B: 37-48). Recently, combinations of a proton pump inhibitors and a single antibiotic have been shown to ameliorate duodenal ulcer disease. (Malfertheiner, P. and J. E. Dominguez-Munoz supra). However, methods employing antibiotic agents can have the  
5 problem of the emergence of bacterial strains which are resistant to these agents. (Hopkins, R. J. and J. G. Morris, supra). These limitations demonstrate that new more effective methods are needed to combat *H. pylori* infections *in vivo*. In particular, the design of new vaccines that may prevent infection by this bacterium is highly desirable.

## 10 Summary of the Invention

This invention relates to novel genes, e.g., genes encoding polypeptides such as bacterial surface proteins, from the organism *Helicobacter pylori* (*H. pylori*), and other related genes, their products, and uses thereof. The nucleic acids and peptides of the present invention have utility for diagnostic and therapeutics for *H. pylori* and other  
15 *Helicobacter* species. They can also be used to detect the presence of *H. pylori* and other *Helicobacter* species in a sample; and for use in screening compounds for the ability to interfere with the *H. pylori* life cycle or to inhibit *H. pylori* infection. More specifically, this invention features compositions of nucleic acids corresponding to entire coding sequences of *H. pylori* proteins, including surface or secreted proteins or  
20 parts thereof, nucleic acids capable of binding mRNA from *H. pylori* proteins to block protein translation, and methods for producing *H. pylori* proteins or parts thereof using peptide synthesis and recombinant DNA techniques. This invention also features antibodies and nucleic acids useful as probes to detect *H. pylori* infection. In addition, vaccine compositions and methods for the protection against infection by *H. pylori* are  
25 within the scope of this invention.

## Detailed Description of the Drawings

Figure 1 is a bar graph that depicts the antibody titer in serum of mice following immunization with specific *H. pylori* antigens.

30 Figure 2 is a bar graph that depicts the antibody titer in mucous of mice following immunization with specific *H. pylori* antigens.

Figure 3 is a bar graph that depicts therapeutic immunization of *H. pylori* infected mice with specific antigens dissolved in HEPES buffer.

Figure 4 is a bar graph that depicts therapeutic immunization of *H. pylori*  
35 infected mice with specific antigens dissolved in buffer containing DOC.

Figure 5 is a graph depicting the activity of recombinant PPIase.

- 3 -

Figure 6 is a graph depicting PPlase activity in an *H. pylori* extract.

Figure 7 is a graph depicting a decrease of glutamate racemase activity with L-Serine-O-Sulfate.

Figure 8 depicts the amino acid sequence alignment in a portion of the sequence  
5 of 12 *H. pylori* proteins (depicted in the single letter amino acid code and designated by their amino acid Sequence ID Numbers; shown N-terminal to C-terminal, left to right).

Figure 9 depicts the N-terminal portion of nine *H. pylori* proteins (depicted in the single letter amino acid code and designated by their amino acid Sequence ID Numbers; shown N-terminal to C-terminal, left to right).

10

#### Detailed Description of the Invention

In one aspect, the invention features a recombinant or substantially pure preparation of *H. pylori* polypeptide of SEQ ID NO: 492. The invention also includes substantially pure nucleic acid encoding an *H. pylori* polypeptide of SEQ ID NO: 492,  
15 such nucleic acid is contained in SEQ ID NO: 1. The *H. pylori* polypeptide sequences described herein are contained in the Sequence Listing, and the nucleic acids encoding *H. pylori* polypeptides are contained in the Sequence Listing.

In another aspect, the invention features a recombinant or substantially pure preparation of an *H. pylori* polypeptide selected from the group consisting of *H. pylori*  
20 polypeptides of SEQ ID NO: 492 through SEQ ID NO: 541. The invention also includes substantially pure nucleic acid encoding an *H. pylori* polypeptide selected from the group consisting of *H. pylori* polypeptides SEQ ID NO: 492 through SEQ ID NO: 541, such nucleic acids are contained in SEQ ID NO: 1 through SEQ ID NO: 50.

In another aspect, the invention features a recombinant or substantially pure preparation of an *H. pylori* polypeptide selected from the group consisting of *H. pylori*  
25 polypeptides of SEQ ID NO: 542 through SEQ ID NO: 591. The invention also includes substantially pure nucleic acid encoding an *H. pylori* polypeptide selected from the group consisting of *H. pylori* polypeptides SEQ ID NO: 542 through SEQ ID NO: 591, such nucleic acids are contained in SEQ ID NO: 51 through SEQ ID NO: 100.

In another aspect, the invention features a recombinant or substantially pure preparation of an *H. pylori* polypeptide selected from the group consisting of *H. pylori*  
30 polypeptides of SEQ ID NO: 592 through SEQ ID NO: 641. The invention also includes substantially pure nucleic acid encoding an *H. pylori* polypeptide selected from the group consisting of *H. pylori* polypeptides SEQ ID NO: 592 through SEQ ID NO: 641, such nucleic acids are contained in SEQ ID NO: 101 through SEQ ID NO: 150.  
35

In another aspect, the invention features a recombinant or substantially pure preparation of an *H. pylori* polypeptide selected from the group consisting of *H. pylori* polypeptides of SEQ ID NO: 642 through SEQ ID NO: 691. The invention also includes substantially pure nucleic acid encoding an *H. pylori* polypeptide selected from the group consisting of *H. pylori* polypeptides SEQ ID NO: 642 through SEQ ID NO: 691, such nucleic acids are contained in SEQ ID NO: 151 through SEQ ID NO: 200.

In another aspect, the invention features a recombinant or substantially pure preparation of an *H. pylori* polypeptide selected from the group consisting of *H. pylori* polypeptides of SEQ ID NO: 692 through SEQ ID NO: 741. The invention also includes substantially pure nucleic acid encoding an *H. pylori* polypeptide selected from the group consisting of *H. pylori* polypeptides SEQ ID NO: 692 through SEQ ID NO: 741, such nucleic acids are contained in SEQ ID NO: 201 through SEQ ID NO: 250.

In another aspect, the invention features a recombinant or substantially pure preparation of an *H. pylori* polypeptide selected from the group consisting of *H. pylori* polypeptides of SEQ ID NO: 742 through SEQ ID NO: 759, SEQ ID NO: 761, SEQ ID NO: 763, SEQ ID NO: 765 through SEQ ID NO: 791. The invention also includes substantially pure nucleic acid encoding an *H. pylori* polypeptide selected from the group consisting of *H. pylori* polypeptides SEQ ID NO: 742 through SEQ ID NO: 759, SEQ ID NO: 761, SEQ ID NO: 763, SEQ ID NO: 765 through SEQ ID NO: 791, such nucleic acids are contained in SEQ ID NO: 251 through SEQ ID NO: 268, SEQ ID NO: 270, SEQ ID NO: 272, and SEQ ID NO: 274 through SEQ ID NO: 300.

In another aspect, the invention features a recombinant or substantially pure preparation of an *H. pylori* polypeptide selected from the group consisting of *H. pylori* polypeptides of SEQ ID NO: 792 through SEQ ID NO: 818 and SEQ ID NO: 820 through SEQ ID NO: 841. The invention also includes substantially pure nucleic acid encoding an *H. pylori* polypeptide selected from the group consisting of *H. pylori* polypeptides SEQ ID NO: 792 through SEQ ID NO: 818 and SEQ ID NO: 820 through SEQ ID NO: 841, such nucleic acids are contained in SEQ ID NO: 301 through SEQ ID NO: 327 and SEQ ID NO: 329 through SEQ ID NO: 350.

In another aspect, the invention features a recombinant or substantially pure preparation of an *H. pylori* polypeptide selected from the group consisting of *H. pylori* polypeptides of SEQ ID NO: 842 through SEQ ID NO: 846 and SEQ ID NO: 848 through SEQ ID NO: 891. The invention also includes substantially pure nucleic acid encoding an *H. pylori* polypeptide selected from the group consisting of *H. pylori* polypeptides SEQ ID NO: 842 through SEQ ID NO: 846 and SEQ ID NO: 848 through

- 5 -

SEQ ID NO: 891, such nucleic acids are contained in SEQ ID NO: 351 through SEQ ID NO: 364 and SEQ ID NO: 366 through SEQ ID NO: 400.

In another aspect, the invention features a recombinant or substantially pure preparation of an *H. pylori* polypeptide selected from the group consisting of *H. pylori* polypeptides of SEQ ID NO: 892 through SEQ ID NO: 896 and SEQ ID NO: 898 through SEQ ID NO: 941. The invention also includes substantially pure nucleic acid encoding an *H. pylori* polypeptide selected from the group consisting of *H. pylori* polypeptides SEQ ID NO: 892 through SEQ ID NO: 896 and SEQ ID NO: 898 through SEQ ID NO: 941, such nucleic acids are contained in SEQ ID NO: 401 through SEQ ID NO: 405 and SEQ ID NO: 407 through SEQ ID NO: 450.

In another aspect, the invention features a recombinant or substantially pure preparation of an *H. pylori* polypeptide selected from the group consisting of *H. pylori* polypeptides of SEQ ID NO: 942 through SEQ ID NO: 963 and SEQ ID NO: 966 through SEQ ID NO: 982. The invention also includes substantially pure nucleic acid encoding an *H. pylori* polypeptide selected from the group consisting of *H. pylori* polypeptides SEQ ID NO: 942 through SEQ ID NO: 963 and SEQ ID NO: 966 through SEQ ID NO: 982, such nucleic acids are contained in SEQ ID NO: 451 through SEQ ID NO: 472 and SEQ ID NO: 475 through SEQ ID NO: 491.

In another aspect, the invention features a recombinant or substantially pure preparation of an *H. pylori* polypeptide selected from the group consisting of *H. pylori* polypeptides of SEQ ID NO: 1037, SEQ ID NO: 1038, SEQ ID NO: 1041 through SEQ ID NO: 1087 and SEQ ID NO: 1090. The invention also includes substantially pure nucleic acid encoding an *H. pylori* polypeptide selected from the group consisting of *H. pylori* polypeptides SEQ ID NO: 1037, SEQ ID NO: 1038, SEQ ID NO: 1041 through SEQ ID NO: 1087 and SEQ ID NO: 1090, such nucleic acids are contained in SEQ ID NO: 983, SEQ ID NO: 984, SEQ ID NO: 987 through SEQ ID NO: 1033 and SEQ ID NO: 1036.

In another aspect, the invention features a recombinant or substantially pure preparation of an *H. pylori* polypeptide selected from the group consisting of *H. pylori* polypeptides of SEQ ID NO: 1296 through SEQ ID NO: 1298. The invention also includes substantially pure nucleic acid encoding an *H. pylori* polypeptide selected from the group consisting of *H. pylori* polypeptides SEQ ID NO: 1296 through SEQ ID NO: 1298, such nucleic acids are contained in SEQ ID NO: 1293 through SEQ ID NO: 1295.

In another aspect, the invention features a recombinant or substantially pure preparation of an *H. pylori* polypeptide selected from the group consisting of *H. pylori* polypeptides as set forth in the Sequence Listing. The invention also includes

substantially pure nucleic acid encoding an *H. pylori* polypeptide selected from the group consisting of *H. pylori* polypeptides as set forth in the Sequence Listing. It should be understood that this invention encompasses each of the *H. pylori* polypeptides and nucleic acids encoding such polypeptides as identified in the Sequence Listing by a  
5 given sequence identification number. For example, a representative *H. pylori* polypeptide is contained in SEQ ID NO: 494. Therefore, this invention encompasses a recombinant or substantially pure preparation of an *H. pylori* polypeptide of SEQ ID NO: 494. The invention also includes substantially pure nucleic acid encoding an *H. pylori* polypeptide of SEQ ID NO: 494.

10 In another aspect, the invention pertains to any individual *H. pylori* polypeptide member or nucleic acid encoding such member from the above-identified groups of *H. pylori* polypeptides (e.g., SEQ ID NO: 542-SEQ ID NO: 591) or nucleic acids (e.g., SEQ ID NO: 51-SEQ ID NO: 100), as well as any subgroups from within the above-identified groups. Furthermore, the subgroups can preferably consists of 1, 3, 5, 10, 15,  
15 20, 30 or 40 members of any of the groups identified above, as well as, any combinations thereof. For example, the group consisting of *H. pylori* polypeptides SEQ ID NO: 692 through SEQ ID NO: 741 can be divided into one or more subgroups as follows: SEQ ID NO: 692-SEQ ID NO: 680; SEQ ID NO: 681-SEQ ID NO: 710; SEQ ID NO: 711-SEQ ID NO: 730; SEQ ID NO: 731-SEQ ID NO: 741; or any combinations  
20 thereof.

Particularly preferred is an isolated nucleic acid comprising a nucleotide sequence encoding an *H. pylori* cell envelope polypeptide or a fragment thereof. Such nucleic acid is selected from the group consisting of SEQ ID NO: 255, SEQ ID NO: 263, SEQ ID NO: 266, SEQ ID NO: 277, SEQ ID NO: 280, SEQ ID NO: 285, SEQ ID  
25 NO: 292, SEQ ID NO: 294, SEQ ID NO: 299, SEQ ID NO: 311, SEQ ID NO: 312, SEQ ID NO: 313, SEQ ID NO: 321, SEQ ID NO: 327, SEQ ID NO: 329, SEQ ID NO: 331, SEQ ID NO: 353, SEQ ID NO: 364, SEQ ID NO: 366, SEQ ID NO: 368, SEQ ID NO: 375, SEQ ID NO: 384, SEQ ID NO: 391, SEQ ID NO: 392, SEQ ID NO: 397, SEQ ID NO: 398, SEQ ID NO: 402, SEQ ID NO: 404, SEQ ID NO: 409, SEQ ID NO: 410, SEQ  
30 ID NO: 412, SEQ ID NO: 427, SEQ ID NO: 433, SEQ ID NO: 434, SEQ ID NO: 441, SEQ ID NO: 444, SEQ ID NO: 445, SEQ ID NO: 449, SEQ ID NO: 450, SEQ ID NO: 452, SEQ ID NO: 453, SEQ ID NO: 466, SEQ ID NO: 468, SEQ ID NO: 469, SEQ ID NO: 983, SEQ ID NO: 989, SEQ ID NO: 1008, SEQ ID NO: 1011, SEQ ID NO: 1014, SEQ ID NO: 1015, SEQ ID NO: 1029, SEQ ID NO: 1032, SEQ ID NO: 259, SEQ ID  
35 NO: 286, SEQ ID NO: 326, SEQ ID NO: 374, SEQ ID NO: 399, SEQ ID NO: 422, SEQ ID NO: 454, SEQ ID NO: 465, SEQ ID NO: 998, SEQ ID NO: 1009, SEQ ID NO:

- 7 -

1023, SEQ ID NO: 1294, SEQ ID NO: 1295, SEQ ID NO: 319, SEQ ID NO: 325, SEQ ID NO: 425, SEQ ID NO: 437, SEQ ID NO: 438, SEQ ID NO: 447, SEQ ID NO: 448, SEQ ID NO: 467, SEQ ID NO: 996, SEQ ID NO: 1027, SEQ ID NO: 1031, SEQ ID NO: 254, SEQ ID NO: 352, SEQ ID NO: 415, SEQ ID NO: 1019, SEQ ID NO: 381, SEQ ID NO: 389, SEQ ID NO: 1010, SEQ ID NO: 1012, SEQ ID NO: 354, SEQ ID NO: 372, SEQ ID NO: 400, SEQ ID NO: 421, SEQ ID NO: 1022, SEQ ID NO: 463, SEQ ID NO: 281, SEQ ID NO: 988, SEQ ID NO: 411, SEQ ID NO: 407, SEQ ID NO: 1017, SEQ ID NO: 290, SEQ ID NO: 417, SEQ ID NO: 430, SEQ ID NO: 992, SEQ ID NO: 1025, SEQ ID NO: 477, SEQ ID NO: 414, SEQ ID NO: 253, SEQ ID NO: 293, SEQ ID NO: 334, SEQ ID NO: 343, SEQ ID NO: 418, SEQ ID NO: 424, and SEQ ID NO: 443.

In another embodiment, the *H. pylori* cell envelope polypeptide or a fragment thereof is an *H. pylori* outer membrane polypeptide or a fragment thereof encoded by the nucleic acid selected from the group consisting of SEQ ID NO: 255, SEQ ID NO: 263, SEQ ID NO: 266, SEQ ID NO: 277, SEQ ID NO: 280, SEQ ID NO: 285, SEQ ID NO: 292, SEQ ID NO: 294, SEQ ID NO: 299, SEQ ID NO: 311, SEQ ID NO: 312, SEQ ID NO: 313, SEQ ID NO: 321, SEQ ID NO: 327, SEQ ID NO: 329, SEQ ID NO: 331, SEQ ID NO: 353, SEQ ID NO: 364, SEQ ID NO: 366, SEQ ID NO: 368, SEQ ID NO: 375, SEQ ID NO: 384, SEQ ID NO: 391, SEQ ID NO: 392, SEQ ID NO: 397, SEQ ID NO: 398, SEQ ID NO: 402, SEQ ID NO: 404, SEQ ID NO: 409, SEQ ID NO: 410, SEQ ID NO: 412, SEQ ID NO: 427, SEQ ID NO: 433, SEQ ID NO: 434, SEQ ID NO: 441, SEQ ID NO: 444, SEQ ID NO: 445, SEQ ID NO: 449, SEQ ID NO: 450, SEQ ID NO: 452, SEQ ID NO: 453, SEQ ID NO: 466, SEQ ID NO: 468, SEQ ID NO: 469, SEQ ID NO: 983, SEQ ID NO: 989, SEQ ID NO: 1008, SEQ ID NO: 1011, SEQ ID NO: 1014, SEQ ID NO: 1015, SEQ ID NO: 1029, SEQ ID NO: 1032, SEQ ID NO: 259, SEQ ID NO: 286, SEQ ID NO: 326, SEQ ID NO: 374, SEQ ID NO: 399, SEQ ID NO: 422, SEQ ID NO: 454, SEQ ID NO: 465, SEQ ID NO: 998, SEQ ID NO: 1009, SEQ ID NO: 1023, SEQ ID NO: 1294, SEQ ID NO: 1295, SEQ ID NO: 319, SEQ ID NO: 325, SEQ ID NO: 425, SEQ ID NO: 437, SEQ ID NO: 438, SEQ ID NO: 447, SEQ ID NO: 448, SEQ ID NO: 467, SEQ ID NO: 996, SEQ ID NO: 1027, SEQ ID NO: 1031, SEQ ID NO: 254, SEQ ID NO: 352, SEQ ID NO: 415, SEQ ID NO: 1019, SEQ ID NO: 381, SEQ ID NO: 389, SEQ ID NO: 1010, and SEQ ID NO: 1012.

In another embodiment, the *H. pylori* outer membrane polypeptide or a fragment thereof is an *H. pylori* polypeptide having a terminal phenylalanine residue or a fragment thereof encoded by the nucleic acid selected from the group consisting of SEQ ID NO: 255, SEQ ID NO: 263, SEQ ID NO: 266, SEQ ID NO: 277, SEQ ID NO: 280,

we

5

10

15

20

25

30

35

- 9 -

In another embodiment, the *H. pylori* inner membrane polypeptide or a fragment thereof is an *H. pylori* polypeptide involved in cofactor metabolism or a fragment thereof encoded by the nucleic acid comprising a nucleotide sequence of SEQ ID NO: 463.

5 In another embodiment, the *H. pylori* inner membrane polypeptide or a fragment thereof is an *H. pylori* polypeptide involved in secretion or adhesion or a fragment thereof encoded by the nucleic acid selected from the group consisting of SEQ ID NO: 281 and SEQ ID NO: 988.

10 In another embodiment, the *H. pylori* inner membrane polypeptide or a fragment thereof is an *H. pylori* polypeptide involved in transport or a fragment thereof encoded by the nucleic acid selected from the group consisting of SEQ ID NO: 407 and SEQ ID NO: 1017.

In another embodiment, the *H. pylori* cell envelope polypeptide or a fragment thereof is an *H. pylori* flagellar polypeptide or a fragment thereof encoded by the nucleic acid comprising a nucleotide sequence of SEQ ID NO: 477.

15 In another embodiment, the *H. pylori* cell envelope polypeptide or a fragment thereof is an *H. pylori* transport polypeptide or a fragment thereof encoded by the nucleic acid comprising a nucleotide sequence of SEQ ID NO: 414.

20 Particularly preferred is an isolated nucleic acid comprising a nucleotide sequence encoding an *H. pylori* cytoplasmic polypeptide or a fragment thereof. Such nucleic acid is selected from the group consisting of SEQ ID NO: 470, SEQ ID NO: 1033, SEQ ID NO: 357, SEQ ID NO: 457, SEQ ID NO: 461, SEQ ID NO: 1030, SEQ ID NO: 345, SEQ ID NO: 383, SEQ ID NO: 387, SEQ ID NO: 455, SEQ ID NO: 1003, SEQ ID NO: 351, SEQ ID NO: 416, SEQ ID NO: 278, SEQ ID NO: 335, SEQ ID NO: 346, SEQ ID NO: 350, SEQ ID NO: 419, SEQ ID NO: 460, SEQ ID NO: 472, SEQ ID NO: 1000, SEQ ID NO: 1004, SEQ ID NO: 1020, SEQ ID NO: 1293, SEQ ID NO: 318, SEQ ID NO: 322, SEQ ID NO: 324, SEQ ID NO: 330, SEQ ID NO: 347, SEQ ID NO: 440, SEQ ID NO: 446, SEQ ID NO: 464, SEQ ID NO: 490, SEQ ID NO: 491, SEQ ID NO: 995, SEQ ID NO: 997, SEQ ID NO: 1005, and SEQ ID NO: 1028.

30 In another embodiment, the *H. pylori* cytoplasmic polypeptide or a fragment thereof is an *H. pylori* polypeptide involved in energy conversion or a fragment thereof encoded by the nucleic acid selected from the group consisting of SEQ ID NO: 470 and SEQ ID NO: 1033.

35 In another embodiment, the *H. pylori* cytoplasmic polypeptide or a fragment thereof is an *H. pylori* polypeptide involved in amino acid metabolism and transport or a fragment thereof encoded by the nucleic acid selected from the group consisting of SEQ ID NO: 357 and SEQ ID NO: 457.

In another embodiment, the *H. pylori* cytoplasmic polypeptide or a fragment thereof is an *H. pylori* polypeptide involved in nucleotide metabolism and transport or a fragment thereof encoded by the nucleic acid selected from the group consisting of SEQ ID NO: 461 and SEQ ID NO: 1030.

5 In another embodiment, the *H. pylori* cytoplasmic polypeptide or a fragment thereof is an *H. pylori* polypeptide involved in cofactor metabolism or a fragment thereof encoded by the nucleic acid selected from the group consisting of SEQ ID NO: 345, SEQ ID NO: 383, SEQ ID NO: 387, SEQ ID NO: 455, and SEQ ID NO: 1003.

10 In another embodiment, the *H. pylori* cytoplasmic polypeptide or a fragment thereof is an *H. pylori* polypeptide involved in lipid metabolism or a fragment thereof encoded by the nucleic acid selected from the group consisting of SEQ ID NO: 351 and SEQ ID NO: 416.

In another embodiment, the *H. pylori* cytoplasmic polypeptide or a fragment thereof is an *H. pylori* polypeptide involved in genome replication, transcription,  
15 recombination and repair or a fragment thereof encoded by the nucleic acid selected from the group consisting of SEQ ID NO: 278, SEQ ID NO: 335, SEQ ID NO: 346, SEQ ID NO: 350, SEQ ID NO: 419, SEQ ID NO: 460, SEQ ID NO: 472, SEQ ID NO: 1000, SEQ ID NO: 1004, SEQ ID NO: 1020, and SEQ ID NO: 1293.

Particularly preferred is an isolated nucleic acid comprising a nucleotide  
20 sequence encoding an *H. pylori* secreted polypeptide or a fragment thereof. Such nucleic acid is selected from the group consisting of SEQ ID NO: 355, SEQ ID NO: 1006, SEQ ID NO: 257, SEQ ID NO: 258, SEQ ID NO: 260, SEQ ID NO: 261, SEQ ID NO: 264, SEQ ID NO: 265, SEQ ID NO: 268, SEQ ID NO: 270, SEQ ID NO: 272, SEQ ID NO: 274, SEQ ID NO: 275, SEQ ID NO: 276, SEQ ID NO: 279, SEQ ID NO: 283,  
25 SEQ ID NO: 284, SEQ ID NO: 287, SEQ ID NO: 288, SEQ ID NO: 289, SEQ ID NO: 291, SEQ ID NO: 295, SEQ ID NO: 296, SEQ ID NO: 297, SEQ ID NO: 298, SEQ ID NO: 300, SEQ ID NO: 301, SEQ ID NO: 302, SEQ ID NO: 303, SEQ ID NO: 304, SEQ ID NO: 305, SEQ ID NO: 314, SEQ ID NO: 315, SEQ ID NO: 323, SEQ ID NO: 338, SEQ ID NO: 342, SEQ ID NO: 348, SEQ ID NO: 349, SEQ ID NO: 356, SEQ ID NO:  
30 358, SEQ ID NO: 359, SEQ ID NO: 360, SEQ ID NO: 361, SEQ ID NO: 362, SEQ ID NO: 363, SEQ ID NO: 367, SEQ ID NO: 370, SEQ ID NO: 371, SEQ ID NO: 373, SEQ ID NO: 377, SEQ ID NO: 378, SEQ ID NO: 379, SEQ ID NO: 380, SEQ ID NO: 388, SEQ ID NO: 390, SEQ ID NO: 394, SEQ ID NO: 395, SEQ ID NO: 396, SEQ ID NO: 401, SEQ ID NO: 403, SEQ ID NO: 405, SEQ ID NO: 408, SEQ ID NO: 420, SEQ ID  
35 NO: 426, SEQ ID NO: 428, SEQ ID NO: 429, SEQ ID NO: 432, SEQ ID NO: 439, SEQ ID NO: 442, SEQ ID NO: 451, SEQ ID NO: 471, SEQ ID NO: 478, SEQ ID NO: 488,

- 11 -

SEQ ID NO: 987, SEQ ID NO: 990, SEQ ID NO: 991, SEQ ID NO: 993, SEQ ID NO: 1001, SEQ ID NO: 1002, SEQ ID NO: 1007, SEQ ID NO: 1013, SEQ ID NO: 1016, SEQ ID NO: 1018, SEQ ID NO: 1021, and SEQ ID NO: 1026.

In another embodiment, the *H. pylori* secreted polypeptide or a fragment thereof is an *H. pylori* polypeptide involved in secretion and adhesion or a fragment thereof encoded by the nucleic acid selected from the group consisting of SEQ ID NO: 355 and SEQ ID NO: 1006.

Particularly preferred is an isolated nucleic acid comprising a nucleotide sequence encoding an *H. pylori* cellular polypeptide or a fragment thereof. Such nucleic acid is selected from the group consisting of SEQ ID NO: 256, SEQ ID NO: 267, SEQ ID NO: 282, SEQ ID NO: 306, SEQ ID NO: 307, SEQ ID NO: 308, SEQ ID NO: 309, SEQ ID NO: 310, SEQ ID NO: 316, SEQ ID NO: 317, SEQ ID NO: 332, SEQ ID NO: 333, SEQ ID NO: 336, SEQ ID NO: 337, SEQ ID NO: 339, SEQ ID NO: 340, SEQ ID NO: 341, SEQ ID NO: 344, SEQ ID NO: 369, SEQ ID NO: 376, SEQ ID NO: 382, SEQ ID NO: 386, SEQ ID NO: 423, SEQ ID NO: 431, SEQ ID NO: 435, SEQ ID NO: 436, SEQ ID NO: 458, SEQ ID NO: 462, SEQ ID NO: 475, SEQ ID NO: 476, SEQ ID NO: 479, SEQ ID NO: 480, SEQ ID NO: 481, SEQ ID NO: 482, SEQ ID NO: 483, SEQ ID NO: 484, SEQ ID NO: 485, SEQ ID NO: 486, SEQ ID NO: 487, SEQ ID NO: 489, SEQ ID NO: 984, SEQ ID NO: 994, SEQ ID NO: 1024, and SEQ ID NO: 1036.

Particularly preferred is a purified or isolated *H. pylori* cell envelope polypeptide or a fragment thereof, wherein the polypeptide is selected from the group consisting of SEQ ID NO: 746, SEQ ID NO: 754, SEQ ID NO: 757, SEQ ID NO: 768, SEQ ID NO: 771, SEQ ID NO: 776, SEQ ID NO: 783, SEQ ID NO: 785, SEQ ID NO: 790, SEQ ID NO: 802, SEQ ID NO: 803, SEQ ID NO: 804, SEQ ID NO: 812, SEQ ID NO: 818, SEQ ID NO: 820, SEQ ID NO: 882, SEQ ID NO: 844, SEQ ID NO: 855, SEQ ID NO: 857, SEQ ID NO: 859, SEQ ID NO: 866, SEQ ID NO: 875, SEQ ID NO: 882, SEQ ID NO: 883, SEQ ID NO: 888, SEQ ID NO: 889, SEQ ID NO: 893, SEQ ID NO: 895, SEQ ID NO: 900, SEQ ID NO: 901, SEQ ID NO: 903, SEQ ID NO: 918, SEQ ID NO: 924, SEQ ID NO: 925, SEQ ID NO: 932, SEQ ID NO: 935, SEQ ID NO: 936, SEQ ID NO: 940, SEQ ID NO: 941, SEQ ID NO: 943, SEQ ID NO: 944, SEQ ID NO: 957, SEQ ID NO: 959, SEQ ID NO: 960, SEQ ID NO: 1037, SEQ ID NO: 1043, SEQ ID NO: 1062, SEQ ID NO: 1065, SEQ ID NO: 1068, SEQ ID NO: 1069, SEQ ID NO: 1083, SEQ ID NO: 1086, SEQ ID NO: 750, SEQ ID NO: 777, SEQ ID NO: 817, SEQ ID NO: 865, SEQ ID NO: 890, SEQ ID NO: 913, SEQ ID NO: 945, SEQ ID NO: 956, SEQ ID NO: 1052, SEQ ID NO: 1063, SEQ ID NO: 1077, SEQ ID NO: 1297, SEQ ID NO: 1298, SEQ ID NO: 810, SEQ ID NO: 816, SEQ ID NO: 916, SEQ ID NO: 928, SEQ ID NO: 929, SEQ

ID NO: 938, SEQ ID NO: 939, SEQ ID NO: 958, SEQ ID NO: 1050, SEQ ID NO: 1081, SEQ ID NO: 1085, SEQ ID NO: 745, SEQ ID NO: 843, SEQ ID NO: 906, SEQ ID NO: 1073, SEQ ID NO: 872, SEQ ID NO: 880, SEQ ID NO: 1064, SEQ ID NO: 1066, SEQ ID NO: 845, SEQ ID NO: 863, SEQ ID NO: 891, SEQ ID NO: 912, SEQ ID NO: 1076, SEQ ID NO: 954, SEQ ID NO: 772, SEQ ID NO: 1042, SEQ ID NO: 902, SEQ ID NO: 898, SEQ ID NO: 1071, SEQ ID NO: 781, SEQ ID NO: 908, SEQ ID NO: 921, SEQ ID NO: 1046, SEQ ID NO: 1079, SEQ ID NO: 968, SEQ ID NO: 905, SEQ ID NO: 744, SEQ ID NO: 784, SEQ ID NO: 825, SEQ ID NO: 834, SEQ ID NO: 909, SEQ ID NO: 915, and SEQ ID NO: 934.

10 In another embodiment, the *H. pylori* cell envelope polypeptide or a fragment thereof is an *H. pylori* outer membrane polypeptide or a fragment thereof selected from the group consisting of SEQ ID NO: 746, SEQ ID NO: 754, SEQ ID NO: 757, SEQ ID NO: 768, SEQ ID NO: 771, SEQ ID NO: 776, SEQ ID NO: 783, SEQ ID NO: 785, SEQ ID NO: 790, SEQ ID NO: 802, SEQ ID NO: 803, SEQ ID NO: 804, SEQ ID NO: 812, 15 SEQ ID NO: 818, SEQ ID NO: 820, SEQ ID NO: 882, SEQ ID NO: 844, SEQ ID NO: 855, SEQ ID NO: 857, SEQ ID NO: 859, SEQ ID NO: 866, SEQ ID NO: 875, SEQ ID NO: 882, SEQ ID NO: 883, SEQ ID NO: 888, SEQ ID NO: 889, SEQ ID NO: 893, SEQ ID NO: 895, SEQ ID NO: 900, SEQ ID NO: 901, SEQ ID NO: 903, SEQ ID NO: 918, SEQ ID NO: 924, SEQ ID NO: 925, SEQ ID NO: 932, SEQ ID NO: 935, SEQ ID NO: 20 936, SEQ ID NO: 940, SEQ ID NO: 941, SEQ ID NO: 943, SEQ ID NO: 944, SEQ ID NO: 957, SEQ ID NO: 959, SEQ ID NO: 960, SEQ ID NO: 1037, SEQ ID NO: 1043, SEQ ID NO: 1062, SEQ ID NO: 1065, SEQ ID NO: 1068, SEQ ID NO: 1069, SEQ ID NO: 1083, SEQ ID NO: 1086, SEQ ID NO: 750, SEQ ID NO: 777, SEQ ID NO: 817, SEQ ID NO: 865, SEQ ID NO: 890, SEQ ID NO: 913, SEQ ID NO: 945, SEQ ID NO: 25 956, SEQ ID NO: 1052, SEQ ID NO: 1063, SEQ ID NO: 1077, SEQ ID NO: 1297, SEQ ID NO: 1298, SEQ ID NO: 810, SEQ ID NO: 816, SEQ ID NO: 916, SEQ ID NO: 928, SEQ ID NO: 929, SEQ ID NO: 938, SEQ ID NO: 939, SEQ ID NO: 958, SEQ ID NO: 1050, SEQ ID NO: 1081, SEQ ID NO: 1085, SEQ ID NO: 745, SEQ ID NO: 843, SEQ ID NO: 906, SEQ ID NO: 1073, SEQ ID NO: 872, SEQ ID NO: 880, SEQ ID NO: 30 1064, and SEQ ID NO: 1066.

In another embodiment, the *H. pylori* outer membrane polypeptide or a fragment thereof is an *H. pylori* polypeptide having a terminal phenylalanine residue or a fragment thereof selected from the group consisting of SEQ ID NO: 746, SEQ ID NO: 754, SEQ ID NO: 757, SEQ ID NO: 768, SEQ ID NO: 771, SEQ ID NO: 776, SEQ ID NO: 783, SEQ ID NO: 785, SEQ ID NO: 790, SEQ ID NO: 802, SEQ ID NO: 803, SEQ ID NO: 804, SEQ ID NO: 812, SEQ ID NO: 818, SEQ ID NO: 820, SEQ ID NO: 882, 35

- 13 -

SEQ ID NO: 844, SEQ ID NO: 855, SEQ ID NO: 857, SEQ ID NO: 859, SEQ ID NO: 866, SEQ ID NO: 875, SEQ ID NO: 882, SEQ ID NO: 883, SEQ ID NO: 888, SEQ ID NO: 889, SEQ ID NO: 893, SEQ ID NO: 895, SEQ ID NO: 900, SEQ ID NO: 901, SEQ ID NO: 903, SEQ ID NO: 918, SEQ ID NO: 924, SEQ ID NO: 925, SEQ ID NO: 932, 5 SEQ ID NO: 935, SEQ ID NO: 936, SEQ ID NO: 940, SEQ ID NO: 941, SEQ ID NO: 943, SEQ ID NO: 944, SEQ ID NO: 957, SEQ ID NO: 959, SEQ ID NO: 960, SEQ ID NO: 1037, SEQ ID NO: 1043, SEQ ID NO: 1062, SEQ ID NO: 1065, SEQ ID NO: 1068, SEQ ID NO: 1069, SEQ ID NO: 1083, and SEQ ID NO: 1086.

10 In another embodiment, the *H. pylori* outer membrane polypeptide or a fragment thereof is an *H. pylori* polypeptide having a C-terminal tyrosine cluster or a fragment thereof selected from the group consisting of SEQ ID NO: 777, SEQ ID NO: 817, SEQ ID NO: 865, SEQ ID NO: 890, SEQ ID NO: 913, SEQ ID NO: 945, SEQ ID NO: 956, SEQ ID NO: 1052, SEQ ID NO: 1063, SEQ ID NO: 1077, SEQ ID NO: 1297, and SEQ ID NO: 1298.

15 In another embodiment, the *H. pylori* outer membrane polypeptide or a fragment thereof is an *H. pylori* polypeptide having a terminal phenylalanine residue and a C-terminal tyrosine cluster or a fragment thereof selected from the group consisting of SEQ ID NO: 810, SEQ ID NO: 816, SEQ ID NO: 916, SEQ ID NO: 928, SEQ ID NO: 929, SEQ ID NO: 938, SEQ ID NO: 939, SEQ ID NO: 958, SEQ ID NO: 1050, SEQ ID NO: 1081, and SEQ ID NO: 1085. 20

In another embodiment, the *H. pylori* cell envelope polypeptide or a fragment thereof is an *H. pylori* inner membrane polypeptide or a fragment thereof selected from the group consisting of SEQ ID NO: 845, SEQ ID NO: 863, SEQ ID NO: 891, SEQ ID NO: 912, SEQ ID NO: 1076, SEQ ID NO: 954, SEQ ID NO: 772, SEQ ID NO: 1042, 25 SEQ ID NO: 902, SEQ ID NO: 898, SEQ ID NO: 1071, SEQ ID NO: 781, SEQ ID NO: 908, SEQ ID NO: 921, SEQ ID NO: 1046, and SEQ ID NO: 1079.

In another embodiment, the *H. pylori* inner membrane polypeptide or a fragment thereof is an *H. pylori* polypeptide involved in outer membrane and cell wall synthesis or a fragment thereof comprising an amino acid sequence of SEQ ID NO: 845.

30 In another embodiment, the *H. pylori* inner membrane polypeptide or a fragment thereof is an *H. pylori* polypeptide involved in energy conversion or a fragment thereof selected from the group consisting of SEQ ID NO: 863, SEQ ID NO: 891, SEQ ID NO: 912, and SEQ ID NO: 1076.

In another embodiment, the *H. pylori* inner membrane polypeptide or a fragment 35 thereof is an *H. pylori* polypeptide involved in cofactor metabolism or a fragment thereof comprising an amino acid sequence of SEQ ID NO: 954.

In another embodiment, the *H. pylori* inner membrane polypeptide or a fragment thereof is an *H. pylori* polypeptide involved in secretion or adhesion or a fragment thereof selected from the group consisting of SEQ ID NO: 772 and SEQ ID NO: 1042.

5 In another embodiment, the *H. pylori* inner membrane polypeptide or a fragment thereof is an *H. pylori* polypeptide involved in transport or a fragment thereof selected from the group consisting of SEQ ID NO: 898 and SEQ ID NO: 1071.

In another embodiment, the *H. pylori* cell envelope polypeptide or a fragment thereof is an *H. pylori* flagellar polypeptide or a fragment thereof comprising an amino acid sequence of SEQ ID NO: 968.

10 In another embodiment, the *H. pylori* cell envelope polypeptide or a fragment thereof is an *H. pylori* transport polypeptide or a fragment thereof comprising an amino acid sequence of SEQ ID NO: 905.

Particularly preferred is a purified or isolated *H. pylori* cytoplasmic polypeptide or a fragment thereof, wherein the polypeptide is selected from the group consisting of  
15 SEQ ID NO: 961, SEQ ID NO: 1087, SEQ ID NO: 848, SEQ ID NO: 948, SEQ ID NO: 952, SEQ ID NO: 1084, SEQ ID NO: 836, SEQ ID NO: 874, SEQ ID NO: 878, SEQ ID NO: 946, SEQ ID NO: 1057, SEQ ID NO: 842, SEQ ID NO: 907, SEQ ID NO: 769, SEQ ID NO: 826, SEQ ID NO: 837, SEQ ID NO: 841, SEQ ID NO: 910, SEQ ID NO: 951, SEQ ID NO: 963, SEQ ID NO: 1054, SEQ ID NO: 1058, SEQ ID NO: 1074, SEQ  
20 ID NO: 1296, SEQ ID NO: 809, SEQ ID NO: 813, SEQ ID NO: 815, SEQ ID NO: 821, SEQ ID NO: 838, SEQ ID NO: 931, SEQ ID NO: 937, SEQ ID NO: 955, SEQ ID NO: 981, SEQ ID NO: 982, SEQ ID NO: 1049, SEQ ID NO: 1051, SEQ ID NO: 1059, and SEQ ID NO: 1082.

25 In another embodiment, the *H. pylori* cytoplasmic polypeptide or a fragment thereof is an *H. pylori* polypeptide involved in energy conversion or a fragment thereof selected from the group consisting of SEQ ID NO: 961 and SEQ ID NO: 1087.

In another embodiment, the *H. pylori* cytoplasmic polypeptide or a fragment thereof is an *H. pylori* polypeptide involved in amino acid metabolism and transport or a fragment thereof selected from the group consisting of SEQ ID NO: 848 and SEQ ID  
30 NO: 948.

In another embodiment, the *H. pylori* cytoplasmic polypeptide or a fragment thereof is an *H. pylori* polypeptide involved in nucleotide metabolism and transport or a fragment thereof selected from the group consisting of SEQ ID NO: 952 and SEQ ID  
NO: 1084.

35 In another embodiment, the *H. pylori* cytoplasmic polypeptide or a fragment thereof is an *H. pylori* polypeptide involved in cofactor metabolism or a fragment

- 15 -

thereof selected from the group consisting of SEQ ID NO: 836, SEQ ID NO: 874, SEQ ID NO: 878, SEQ ID NO: 946, and SEQ ID NO: 1057.

In another embodiment, the *H. pylori* cytoplasmic polypeptide or a fragment thereof is an *H. pylori* polypeptide involved in lipid metabolism or a fragment thereof  
5 selected from the group consisting of SEQ ID NO: 842, SEQ ID NO: 907.

In another embodiment, the *H. pylori* cytoplasmic polypeptide or a fragment thereof is an *H. pylori* polypeptide involved in genome replication, transcription, recombination and repair or a fragment thereof selected from the group consisting of  
10 SEQ ID NO: 769, SEQ ID NO: 826, SEQ ID NO: 837, SEQ ID NO: 841, SEQ ID NO: 910, SEQ ID NO: 951, SEQ ID NO: 963, SEQ ID NO: 1054, SEQ ID NO: 1058, SEQ ID NO: 1074, and SEQ ID NO: 1296.

Particularly preferred is a purified or isolated *H. pylori* secreted polypeptide or a fragment thereof, wherein the polypeptide is selected from the group consisting of SEQ ID NO: 846, SEQ ID NO: 1060, SEQ ID NO: 748, SEQ ID NO: 749, SEQ ID NO: 751,  
15 SEQ ID NO: 752, SEQ ID NO: 755, SEQ ID NO: 756, SEQ ID NO: 759, SEQ ID NO: 761, SEQ ID NO: 763, SEQ ID NO: 765, SEQ ID NO: 766, SEQ ID NO: 767, SEQ ID NO: 770, SEQ ID NO: 774, SEQ ID NO: 775, SEQ ID NO: 778, SEQ ID NO: 779, SEQ ID NO: 780, SEQ ID NO: 782, SEQ ID NO: 786, SEQ ID NO: 787, SEQ ID NO: 788, SEQ ID NO: 789, SEQ ID NO: 791, SEQ ID NO: 792, SEQ ID NO: 793, SEQ ID NO:  
20 794, SEQ ID NO: 795, SEQ ID NO: 796, SEQ ID NO: 805, SEQ ID NO: 806, SEQ ID NO: 814, SEQ ID NO: 829, SEQ ID NO: 833, SEQ ID NO: 839, SEQ ID NO: 840, SEQ ID NO: 847, SEQ ID NO: 849, SEQ ID NO: 850, SEQ ID NO: 851, SEQ ID NO: 852, SEQ ID NO: 853, SEQ ID NO: 854, SEQ ID NO: 858, SEQ ID NO: 861, SEQ ID NO: 862, SEQ ID NO: 864, SEQ ID NO: 868, SEQ ID NO: 869, SEQ ID NO: 870, SEQ ID  
25 NO: 871, SEQ ID NO: 879, SEQ ID NO: 881, SEQ ID NO: 885, SEQ ID NO: 886, SEQ ID NO: 887, SEQ ID NO: 892, SEQ ID NO: 894, SEQ ID NO: 896, SEQ ID NO: 899, SEQ ID NO: 911, SEQ ID NO: 917, SEQ ID NO: 919, SEQ ID NO: 920, SEQ ID NO: 923, SEQ ID NO: 930, SEQ ID NO: 933, SEQ ID NO: 942, SEQ ID NO: 962, SEQ ID NO: 969, SEQ ID NO: 979, SEQ ID NO: 1041, SEQ ID NO: 1044, SEQ ID NO: 1045,  
30 SEQ ID NO: 1047, SEQ ID NO: 1055, SEQ ID NO: 1056, SEQ ID NO: 1061, SEQ ID NO: 1067, SEQ ID NO: 1070, SEQ ID NO: 1072, SEQ ID NO: 1075, and SEQ ID NO: 1080.

In another embodiment, the *H. pylori* secreted polypeptide or a fragment thereof is an *H. pylori* polypeptide involved in secretion and adhesion or a fragment thereof  
35 selected from the group consisting of SEQ ID NO: 846 and SEQ ID NO: 1060.

Particularly preferred is a purified or isolated *H. pylori* cellular polypeptide or a fragment thereof, wherein the polypeptide is selected from the group consisting of SEQ ID NO: 747, SEQ ID NO: 758, SEQ ID NO: 773, SEQ ID NO: 797, SEQ ID NO: 798, SEQ ID NO: 799, SEQ ID NO: 800, SEQ ID NO: 801, SEQ ID NO: 807, SEQ ID NO: 808, SEQ ID NO: 823, SEQ ID NO: 824, SEQ ID NO: 827, SEQ ID NO: 828, SEQ ID NO: 830, SEQ ID NO: 831, SEQ ID NO: 832, SEQ ID NO: 835, SEQ ID NO: 860, SEQ ID NO: 867, SEQ ID NO: 873, SEQ ID NO: 877, SEQ ID NO: 914, SEQ ID NO: 922, SEQ ID NO: 926, SEQ ID NO: 927, SEQ ID NO: 949, SEQ ID NO: 953, SEQ ID NO: 966, SEQ ID NO: 967, SEQ ID NO: 970, SEQ ID NO: 971, SEQ ID NO: 972, SEQ ID NO: 973, SEQ ID NO: 974, SEQ ID NO: 975, SEQ ID NO: 976, SEQ ID NO: 977, SEQ ID NO: 978, SEQ ID NO: 980, SEQ ID NO: 1038, SEQ ID NO: 1048, SEQ ID NO: 1078, and SEQ ID NO: 1090.

In another aspect, the invention pertains to any individual *H. pylori* polypeptide member or nucleic acid encoding such a member from the above-identified groups of *H. pylori* polypeptides.

In another aspect, the invention features nucleic acids capable of binding mRNA of *H. pylori*. Such nucleic acid is capable of acting as antisense nucleic acid to control the translation of mRNA of *H. pylori*. A further aspect features a nucleic acid which is capable of binding specifically to an *H. pylori* nucleic acid. These nucleic acids are also referred to herein as complements and have utility as probes and as capture reagents.

In another aspect, the invention features an expression system comprising an open reading frame corresponding to *H. pylori* nucleic acid. The nucleic acid further comprises a control sequence compatible with an intended host. The expression system is useful for making polypeptides corresponding to *H. pylori* nucleic acid.

In another aspect, the invention features a cell transformed with the expression system to produce *H. pylori* polypeptides.

In another aspect, the invention features a method of generating antibodies against *H. pylori* polypeptides which are capable of binding specifically to *H. pylori* polypeptides. Such antibody has utility as reagents for immunoassays to evaluate the abundance and distribution of *H. pylori*-specific antigens.

In another aspect, the invention features a method of generating vaccines for immunizing an individual against *H. pylori*. The method includes: immunizing a subject with an *H. pylori* polypeptide, e.g., a surface or secreted polypeptide, or active portion thereof, and a pharmaceutically acceptable carrier. Such vaccines have therapeutic and prophylactic utilities.

- 17 -

In another aspect, the invention provides a method for generating a vaccine comprising a modified immunogenic *H. pylori* polypeptide, e.g., a surface or secreted polypeptide, or active portion thereof, and a pharmacologically acceptable carrier.

5 In another aspect, the invention features a method of evaluating a compound, e.g. a polypeptide, e.g., a fragment of a host cell polypeptide, for the ability to bind an *H. pylori* polypeptide. The method includes: contacting the candidate compound with an *H. pylori* polypeptides and determining if the compound binds or otherwise interacts with an *H. pylori* polypeptide. Compounds which bind *H. pylori* are candidates as activators or inhibitors of the bacterial life cycle. These assays can be performed *in vitro* or *in vivo*.  
10

In another aspect, the invention features a method of evaluating a compound, e.g. a polypeptide, e.g., a fragment of a host cell polypeptide, for the ability to bind an *H. pylori* nucleic acid, e.g., DNA or RNA. The method includes: contacting the candidate compound with an *H. pylori* nucleic acid and determining if the compound binds or  
15 otherwise interacts with an *H. pylori* polypeptide. Compounds which bind *H. pylori* are candidates as activators or inhibitors of the bacterial life cycle. These assays can be performed *in vitro* or *in vivo*.

The invention features *H. pylori* polypeptides, preferably a substantially pure preparation of an *H. pylori* polypeptide, or a recombinant *H. pylori* polypeptide. In preferred embodiments: the polypeptide has biological activity; the polypeptide has an amino acid sequence at least 60%, 70%, 80%, 90%, 95%, 98%, or 99% homologous to  
20 an amino acid sequence contained in the Sequence Listing; the polypeptide has an amino acid sequence essentially the same as an amino acid sequence in the Sequence Listing; the polypeptide is at least 5, 10, 20, 50, 100, or 150 amino acid residues in length; the  
25 polypeptide includes at least 5, preferably at least 10, more preferably at least 20, more preferably at least 50, 100, or 150 contiguous amino acid residues of a polypeptide contained in the Sequence Listing.

In preferred embodiments: the *H. pylori* polypeptide is encoded by a nucleic acid contained in the Sequence Listing, or by a nucleic acid having at least 60%, 70%,  
30 80%, 90%, 95%, 98%, or 99% homology with a nucleic acid shown in the Sequence Listing.

In a preferred embodiment, the subject *H. pylori* polypeptide differs in amino acid sequence at 1, 2, 3, 5, 10 or more residues from a sequence in the Sequence Listing. The differences, however, are such that the *H. pylori* polypeptide exhibits an *H. pylori*  
35 biological activity, e.g., the *H. pylori* polypeptide retains a biological activity of a naturally occurring *H. pylori* enzyme.

In preferred embodiments, the polypeptide includes all or a fragment of an amino acid sequence contained in the Sequence Listing; fused, in reading frame, to additional amino acid residues, preferably to residues encoded by genomic DNA 5' to the genomic DNA which encodes a sequence contained in the Sequence Listing.

5 In yet other preferred embodiments, the *H. pylori* polypeptide is a recombinant fusion protein having a first *H. pylori* polypeptide portion and a second polypeptide portion, e.g., a second polypeptide portion having an amino acid sequence unrelated to *H. pylori*. The second polypeptide portion can be, e.g., any of glutathione-S-transferase, a DNA binding domain, or a polymerase activating domain. In preferred embodiment  
10 the fusion protein can be used in a two-hybrid assay.

Polypeptides of the invention include those which arise as a result of alternative transcription events, alternative RNA splicing events, and alternative translational and postranslational events.

The invention also encompasses an immunogenic component which includes an  
15 *H. pylori* polypeptide in an immunogenic preparation; the immunogenic component being capable of eliciting an immune response specific for the *H. pylori* polypeptide, e.g., a humoral response, an antibody response, or a cellular response. In preferred embodiments, the immunogenic component comprises at least one antigenic determinant from a polypeptide contained in the Sequence Listing.

20 In another aspect, the invention provides a substantially pure nucleic acid having a nucleotide sequence which encodes an *H. pylori* polypeptide. In preferred embodiments: the encoded polypeptide has biological activity; the encoded polypeptide has an amino acid sequence at least 60%, 70%, 80%, 90%, 95%, 98%, or 99% homologous to an amino acid sequence contained in the Sequence Listing; the encoded  
25 polypeptide has an amino acid sequence essentially the same as an amino acid sequence in the Sequence Listing; the encoded polypeptide is at least 5, 10, 20, 50, 100, or 150 amino acids in length; the encoded polypeptide comprises at least 5, preferably at least 10, more preferably at least 20, more preferably at least 50, 100, or 150 contiguous amino acids contained in the Sequence Listing.

30 In preferred embodiments: the nucleic acid is that as shown in the Sequence Listing; the nucleic acid is at least 60%, 70%, 80%, 90%, 95%, 98%, or 99% homologous with a nucleic acid sequence contained in the Sequence Listing.

In a preferred embodiment, the encoded *H. pylori* polypeptide differs (e.g., by amino acid substitution, addition or deletion of at least one amino acid residue) in amino  
35 acid sequence at 1, 2, 3, 5, 10 or more residues, from a sequence in the Sequence Listing. The differences, however, are such that: the *H. pylori* encoded polypeptide

- 19 -

exhibits a *H. pylori* biological activity, e.g., the encoded *H. pylori* enzyme retains a biological activity of a naturally occurring *H. pylori*.

In preferred embodiments, the encoded polypeptide includes all or a fragment of an amino acid sequence contained in the Sequence Listing; fused, in reading frame, to  
5 additional amino acid residues, preferably to residues encoded by genomic DNA 5' to the genomic DNA which encodes a sequence contained in the Sequence Listing.

In preferred embodiments, the subject *H. pylori* nucleic acid will include a transcriptional regulatory sequence, e.g. at least one of a transcriptional promoter or transcriptional enhancer sequence, operably linked to the *H. pylori* gene sequence, e.g.,  
10 to render the *H. pylori* gene sequence suitable for expression in a recombinant host cell.

In yet a further preferred embodiment, the nucleic acid which encodes an *H. pylori* polypeptide of the invention, hybridizes under stringent conditions to a nucleic acid probe corresponding to at least 8 consecutive nucleotides of a nucleic acid contained in the Sequence Listing; more preferably to at least 12 consecutive nucleotides  
15 of a nucleic acid contained in the Sequence Listing; more preferably to at least 20 consecutive nucleotides of a nucleic acid contained in the Sequence Listing; more preferably to at least 40 consecutive nucleotides of a nucleic acid contained in the Sequence Listing.

In a preferred embodiment, the nucleic acid encodes a peptide which differs by at  
20 least one amino acid residue from the sequences shown in the Sequence Listing.

In a preferred embodiment, the nucleic acid differs by at least one nucleotide from a nucleotide sequence shown in the Sequence Listing which encodes amino acids shown in the Sequence Listing.

In another aspect, the invention encompasses: a vector including a nucleic acid  
25 which encodes an *H. pylori* polypeptide or an *H. pylori* polypeptide variant as described herein; a host cell transfected with the vector; and a method of producing a recombinant *H. pylori* polypeptide or *H. pylori* polypeptide variant; including culturing the cell, e.g., in a cell culture medium, and isolating the *H. pylori* or *H. pylori* polypeptide variant e.g., from the cell or from the cell culture medium.

30 In another aspect, the invention features, a purified recombinant nucleic acid having at least 50%, 60%, 70%, 80%, 90%, 95%, 98%, or 99% homology with a sequence contained in the Sequence Listing.

The invention also provides a probe or primer which includes a substantially purified oligonucleotide. The oligonucleotide includes a region of nucleotide sequence  
35 which hybridizes under stringent conditions to at least 10 consecutive nucleotides of sense or antisense sequence contained in the Sequence Listing, or naturally occurring

mutants thereof. In preferred embodiments, the probe or primer further includes a label group attached thereto. The label group can be, e.g., a radioisotope, a fluorescent compound, an enzyme, and/or an enzyme co-factor. Preferably, the oligonucleotide is at least 10 and less than 20, 30, 50, 100, or 150 nucleotides in length.

5 The invention further provides nucleic acids, e.g., RNA or DNA, encoding a polypeptide of the invention. This includes double stranded nucleic acids as well as coding and antisense single strands.

The *H. pylori* strain, from which genomic sequences have been sequenced, has been deposited in the American Type Culture Collection(ATCC) as strain HP-J99.

10 Included in the invention are: allelic variations; natural mutants; induced mutants; proteins encoded by DNA that hybridizes under high or low stringency conditions to a nucleic acid which encodes a polypeptide as shown in the Sequence Listing (for definitions of high and low stringency see Current Protocols in Molecular Biology, John Wiley & Sons, New York, 1989, 6.3.1 - 6.3.6, hereby incorporated by  
15 reference); and, polypeptides specifically bound by antisera to *H. pylori* polypeptides, especially by antisera to an active site or binding domain of *H. pylori* polypeptide. The invention also includes fragments, preferably biologically active fragments. These and other polypeptides are also referred to herein as *H. pylori* polypeptide analogs or variants.

20 Putative functions have been determined for several of the *H. pylori* polypeptides of the invention, as shown in Table 1.

Accordingly, uses of the claimed *H. pylori* polypeptides in these identified functions are also within the scope of the invention.

In addition, the present invention encompasses *H. pylori* polypeptides  
25 characterized as shown in Table 1 below, including: *H. pylori* cell envelope proteins, *H. pylori* secreted proteins, *H. pylori* cytoplasmic proteins and *H. pylori* cellular proteins. Members of these groups were identified by BLAST homology searches and by searches for secretion signal or transmembrane protein motifs. Polypeptides related by  
30 significant homology to the polypeptides of Table 1 (as reflected by FASTA comparisons of the amino acid sequences and indicated in many cases in Tables 3-6 below) are also considered to be classified in the manner of the homolog shown in Table 1.

TABLE 1

ORF NAME	nt Seq ID#	aa Seq ID #
A.CELL ENVELOPE		
A.1 Outer membrane		
A.1.2 Terminal phe residue		
01ae12001_24218781_f2_18	255	746

- 21 -

02ge11622_875260_f3_36	263	754
02gp20706_23632775_f3_32	266	757
07ap11015_23938312_f3_2	271	762
05ee10816_4103408_f2_11	277	768
06ap20306_23437632_f3_9	280	771
07ap20216_7227202_f3_10	285	776
14ap10815_20585777_c1_13	292	783
hp1pl14013_11726503_c2_20	294	785
hp5pl15641_21698387_c2_20	299	790
02gp20706_16803513_f1_1	311	802
02gp20706_20365905_f2_8	312	803
02gp20814_3984818_f1_1	313	804
05ae30220_9882767_f2_34	321	812
07ep11916_5913592_f3_18	327	818
09ze10333_22460750_f2_6	329	820
14ap10221_13689381_c3_4	331	882
06ce20610_1367157_f1_8.aa	353	844
02ge10116_16803513_f2_34	364	855
02ge10116_36367936_c1_92	366	857
05ce10910_25598277_f3_3	368	859
06cp11217_4881263_f2_9	375	866
06ep30223_34409437_f3_94	384	875
06gp71906_970325_c3_190	391	882
07ae10923_24426508_f1_1	392	883
09cp10224_1062966_c3_61	397	888
09cp10224_1412715_c3_56	398	889
09cp61003_14562637_c2_93	402	893
09cp61003_24063587_c1_74	404	895
11ae80818_7952_c1_49	409	900
11ap20714_4960432_c3_97	410	901
11ap20714_7227202_f3_43.aa	412	903
hp5pl15641_12195281_c1_24	427	918
hp6pl0903_4398263_f3_6.aa	433	924
hp6pl0903_4398263_f3_6.aa	434	925
02ae31010_30208317_f1_14	441	932
02cp10615_26573462_c1_45	444	935
03ae10804_12609533_c1_26	445	936
04ep41903_4101593_f2_10.aa	449	940
05ep10815_26570332_c2_99	450	941
05ep10815_4719175_c1_83	452	943
06cp30603_679218_f2_34	453	944
11ae80818_19632781_c3_57	466	957
11ap20714_34023312_f3_46	468	959
11eel1408_4977193_c1_41.aa	469	960
11ael1922_12586675_f2_1	983	1037
07eel1402_2458267_c3_108	985	1039
11ap20714_7227202_f3_40	989	1043
hp7el10192_25598277_c2_15	1008	1062
06ep30223_34409437_f2_64	1011	1065
09cp10224_1062966_c1_44	1014	1068
01ce61016_12931513_c2_106	1015	1069
05ep10815_4719175_c1_115	1029	1083

- 22 -

05ae30220_4977193_c3_198	1032	1086
A.1.3 No terminal phe residue		
02ae11612_22477267_f2_27	259	750
A.1.4 C-terminal tyrosine cluster motif		
07ep11916_5273452_c3_31	286	777
07ce11019_22051291_f1_1	326	817
06cp11217_19720300_f3_11	374	865
09cp10224_429510_c2_46.aa	399	890
hp4e53394_22864682_c2_86.aa	422	913
06ep10615_961562_f2_41	454	945
09cp61003_492187_c2_80.aa	465	956
14gp11423_26803801_f3_7	998	1052
hp4e53394_19720300_c3_98	1009	1063
hp4e53394_19720300_c3_98	1023	1077
01ce61016_492187_c3_120	1294	1297
06ep10615_961562_f1_15	1295	1298
A.1.5 Terminal phe residue and C-terminal tyrosine cluster		
04ge10816_33726080_c2_29	319	810
06ee10709_21675012_f1_2	325	816
hp5p15575_33445317_f2_20.aa	425	916
01cp20708_36134808_f2_11	437	928
02ae31010_12504512_f3_28.aa	438	929
04ep41903_26757937_f3_16	447	938
04ep41903_26757937_f3_16	448	939
11ae80818_7290627_c2_51	467	958
13ae10610_35912_f2_3	996	1050
01cp20708_13086002_f3_27	1027	1081
14cp11908_24218954_c1_68	1031	1085
A.1.6 Via homolgy		
01ae11010_40688_c2_38	254	745
02gp20814_24415958_f3_9	269	760
09cp11003_5945252_f2_4	328	819
06cp30603_23476568_c2_133.aa	352	843
09cp61003_5945252_f1_5	406	897
14ee41924_23527267_c3_107	415	906
07ee11402_10759567_c2_86	1019	1073
A.1.7 Other outer membrane proteins		
06ep10615_9842_f3_46	381	872
06gp10409_3398427_f2_12	389	880
06ep10615_9842_f1_5	1010	1064
06gp10409_3398427_f2_12	1012	1066
A.2 Inner Membrane		
A.2.1 Proteins involved in outer membrane and cell wall synthesis		
06ep30223_4698838_f2_55	354	845
A.2.2 Proteins involved in energy conversion		
06ce20610_4331338_f3_18	372	863
09cp10224_4484718_c1_38	400	891
hp4e13394_5964452_c2_97	421	912
hp4e13394_15828963_c2_90	1022	1076
A.2.3 Proteins involved in cofactor metabolism		
06gp71906_25478192_c1_131	463	954

- 23 -

A.2.4 Proteins involved in secretion and adhesion		
06cp30603_23452_c3_80	281	772
09cp10713_23452_c3_195	988	1042
A.2.5 Via homolgy		
11ap20714_5271967_c1_60	411	902
A.2.6 Proteins involved in transport		
11ae80818_11188791_c3_60	407	898
14cp11908_25593768_c3_97	1017	1071
A.2.7 Other inner membrane proteins		
13ae10712_14100018_f2_12	290	781
hp1p13939_25397327_f3_22	417	908
hp5p15870_14350428_f1_1	430	921
06ge20501_14100018_c1_34	992	1046
05ae30220_14350428_f3_91	1025	1079
A.3 Flagellar proteins		
hp4e13394_3368767_c1_80	477	968
A.4 Transport proteins		
14ce31519_15635927_f3_15	414	905
A.5 Other cell envelope proteins		
04cp11202_24256567_c3_117	253	744
29ge30321_34157812_f3_10	293	784
29ge30321_12913562_f1_1	334	825
hp6p10233_12273302_f1_1	343	834
hp2e10911_24855312_c1_69	418	909
hp5p15575_29300311_c1_29	424	915
02ae31010_5085162_c1_47	443	934
B. CYTOPLASMIC PROTEINS		
B.1 Proteins involved in energy conversion		
11ge10308_5256_f2_1	470	961
11ge10308_24609417_f2_1	1033	1087
B.2 Proteins involved in amino acid metabolism and transport		
06ep11917_24803153_c3_24	357	848
06ep11202_4884677_c1_17	457	948
B.3 Proteins involved in nucleotide metabolism and transport		
06ep30223_23476067_c1_119	461	952
06ep30223_23476067_c1_115	1030	1084
B.4 Proteins involved in cofactor metabolism		
07ap11213_35156577_c1_24	345	836
06ep30223_23557202_c2_130	383	874
06ep30223_5109443_c1_109	387	878
06ep11202_133293_c1_19	455	946
07ee50709_35156577_f3_80	1003	1057
B.5 Proteins involved in lipid metabolism		
hp6e12267_14650278_f3_29	351	842
14ee41924_23834800_f2_32	416	907
B.6 Proteins involved in mRNA translation and ribosome biogenesis		
14ee41924_16282067_c1_72	473	964
07ee11402_19565702_c2_88	1034	1088
" B.7 Proteins involved in genome replication, transcription, recombination and repair"		
05ee10816_4687651_c1_22	278	769

29ge30321_135253_f2_6	335	826
07ap80601_976413_f3_9	346	837
14ce21516_85786_f1_1	350	841
hp2e10911_3349_c1_63	419	910
06ep30223_16512_c3_160	460	951
14ce61516_13073577_f2_12	472	963
07ee50709_4818967_f2_43	1000	1054
05ae30220_976413_c3_204	1004	1058
07ee50709_4818967_f2_43	1020	1074
hp7e10590_13073577_c3_107	1293	1296
B.8 Other cytoplasmic proteins		
04ge10816_22086531_f2_10	318	809
05cp21223_4725443_f3_14	322	813
06ap11119_24426508_f3_26	324	815
12ge10321_4821082_f3_14	330	821
hp3p10807_189075_f2_4	347	838
02ae31010_2117087_f3_34	440	931
03ae10804_21698400_c2_32	446	937
06gp71906_25504187_f3_112	464	955
hp7p10290_35156558_f3_15	490	981
hp7p10290_4351718_f1_6	491	982
13ae10610_859692_c2_32	995	1049
06ap11119_24426508_f3_27	997	1051
hp3e11188_47327_f2_9	1005	1059
07ee50709_26438968_f2_36	1028	1082
C. SECRETED PROTEINS		
C.1 Proteins involved in secretion and adhesion		
12ap10324_4805318_f2_3	355	846
12ap10324_4805318_f2_6	1006	1060
C.2 Other secreted proteins		
01gp11016_4103403_c2_13	257	748
02ael1612_1074212_f1_1	258	749
02ael1612_23598175_f1_2	260	751
02ael1612_33203250_c1_51	261	752
02gp20706_1203402_c3_58	264	755
02gp20706_15781452_c2_51	265	756
02gp20706_4892558_f3_19	268	759
04cp11202_24261588_f2_23	270	761
05ae30220_21619067_f3_56	272	763
07ap11213_35401528_c1_21	273	764
05ae30220_24882812_c3_103	274	765
05ae30220_25953163_c3_98	275	766
05ee10816_14649077_f3_18	276	767
06ap11119_16594193_f1_9	279	770
06cp30603_4689068_c3_79	283	774
07ap11111_234693_c3_14	284	775
09cp11003_19532625_c3_17	287	778
09cp20502_24001388_c1_31	288	779
12gp31106_3024126_f2_25	289	780
13ae10712_29569208_c2_27	291	782
hp2p10272_23697200_f3_22	295	786
hp2p10272_26829136_f1_1	296	787

- 25 -

hp5e15211_819455_c2_24	297	788
hp5p15212_34064750_f2_9	298	789
hp6e10967_23476502_f2_6	300	791
hp6e10967_24882750_f2_7	301	792
hp6e12267_4876718_f2_23	302	793
hp6e20339_1190660_c2_46	303	794
hp6e20339_21492187_c1_40	304	795
hp6e20339_34024187_c1_37	305	796
04cp11202_16603425_c2_72	314	805
04cp11202_19797128_f1_5	315	806
05ee10816_259703_f2_7	323	814
hp2p10272_22692325_f3_21	338	829
hp6e20339_24317062_c3_57	342	833
02ge11622_21695936_c1_54	348	839
12ge10321_24308513_f3_20	349	840
14ee41924_2458267_c2_93	356	847
01ce11104_36125337_c1_8	358	849
01ce21104_33203250_c3_87	359	850
02ae31010_34616666_f2_27	360	851
02ae31010_35270000_f3_33	361	852
02ae31010_36132785_f2_29	362	853
02ge10116_15781452_c1_87	363	854
03ae10804_23485968_c3_47	367	858
06ce20610_29298537_c2_32	370	861
06ce20610_3913967_c3_36	371	862
06cp11118_212827_c1_17	373	864
06cp30603_21492187_f2_41	377	868
06cp30603_34024187_f1_20	378	869
06cp30603_34024187_f1_20	379	870
06ep10615_14649077_f3_52	380	871
06ep30223_5271902_c1_106	388	879
06gp71906_24261588_c2_174	390	881
09ce10413_414011_f1_3	394	885
09ce10413_5865665_f1_4	395	886
09ce52017_29324062_c1_21	396	887
09cp21607_7224187_c2_12	401	892
09cp61003_19532625_c1_78	403	894
09cp61003_24335762_c3_111	405	896
11ae80818_783127_c3_63	408	899
hp4e13394_35957200_f1_21	420	911
hp5p15575_6140713_f2_18	426	917
hp5p15641_24304527_c3_35	428	919
hp5p15641_25635452_c3_34	429	920
hp6p10606_19546933_c3_31	432	923
02ae31010_16833312_f2_19	439	930
02ae31010_36132785_f2_29	442	933
05ep10815_4195292_c1_84	451	942
12ap10324_13178562_f3_6	471	962
hp2e10911_4882027_c2_87	474	965
hp5p15212_6928132_c3_34	478	969
hp7p10290_25548812_f3_14	488	979
07ee50709_10213593_f3_77	986	1040

- 26 -

06ep10615_14649077_f2_30	987	1041
01ce61016_23609580_c3_139	990	1044
06gp71906_3024126_c1_128	991	1045
09cp10713_34024187_f1_31	993	1047
02ap11117_23495187_c3_81	1001	1055
09cp10713_34024187_f1_31	1002	1056
hpl1e80523_23485968_c2_49	1007	1061
09ce10413_5865665_f1_4	1013	1067
01ce61016_23609580_c3_139	1016	1070
14cp11908_783127_c1_72	1018	1072
hp4e13394_5088562_f3_54	1021	1075
hp8e10080_19546933_c2_88	1026	1080
07ee50709_960952_f2_47	1035	1089
D. OTHER CELLULAR PROTEINS		
01ge11619_23711062_c3_14	256	747
02gp20706_23866562_c2_53	267	758
06cp30603_23476568_c1_44	282	773
01ge10203_35281542_c3_16	306	797
01ge10203_860166_f3_9	307	798
01ge11619_13788141_c2_11	308	799
01ge11619_24415880_c2_12	309	800
01ge11619_24417813_c1_8	310	801
04cp11202_23553177_c1_75	316	807
04cp11202_23553177_c3_109	317	808
29ep10720_24220926_f2_8	332	823
29ep10720_24432762_c3_39	333	824
29ge30321_21673965_f2_7	336	827
29ge30321_24336712_f1_5	337	828
hp2p10272_24406280_c1_26	339	830
hp3p10807_29343768_f1_1	340	831
hp3p10807_29352212_f2_5	341	832
02ep20506_24611325_f2_6	344	835
06ae11016_30579712_f2_21	369	860
06cp11217_4897077_f1_6	376	867
06ep11202_26353438_c1_22	382	873
06ep30223_4876077_c3_149	386	877
hp5e15044_4554652_f3_3	423	914
hp6p10590_23440913_c2_31	431	922
hp6p10904_2214676_c1_14	435	926
hp6p10904_23704412_f2_5	436	927
06ep11202_792962_c2_26.aa	458	949
06gp71906_15115637_f2_59	462	953
hp3e11188_47327_f2_5	475	966
hp3e11188_5082842_f3_12	476	967
hp5p15641_30273312_c2_28	479	970
hp5p15641_5211687_c2_29	480	971
hp6p10590_30521093_f2_14	481	972
hp6p10904_7089062_c1_16	482	973
hp6p12129_16603417_f3_14	483	974
hp6p12244_3948467_c1_52	484	975
hp6p22217_23470967_f1_4	485	976
hp7e10192_4412568_f2_5	486	977

- 27 -

hp7p10287_24611325_c2_24	487	978
hp7p10290_25585941_f3_12	489	980
02ge10116_23866562_c3_146	984	1038
hp4p62853_5914693_c3_52	994	1048
07ce10312_4554652_f3_2	1024	1078
hp6p12244_3948467_c3_88	1036	1090

[In Table 1, "nt" represents nucleotide Seq. ID number and "aa" represents amino acid Seq. ID number]

## 5 Definitions

A purified preparation or a substantially pure preparation of a polypeptide, as used herein, means a polypeptide that has been separated from other proteins, lipids, and nucleic acids with which it naturally occurs. Preferably, the polypeptide is also separated from substances, e.g., antibodies or gel matrix, e.g., polyacrylamide, which are used to purify it. Preferably, the polypeptide constitutes at least 10, 20, 50 70, 80 or 95% dry weight of the purified preparation. Preferably, the preparation contains: sufficient polypeptide to allow protein sequencing; at least 1, 10, or 100 µg of the polypeptide; at least 1, 10, or 100 mg of the polypeptide.

A purified preparation of cells refers to, in the case of plant or animal cells, an *in vitro* preparation of cells and not an entire intact plant or animal. In the case of cultured cells or microbial cells, it consists of a preparation of at least 10% and more preferably 50% of the subject cells.

A substantially pure nucleic acid, e.g., a substantially pure DNA, is a nucleic acid which is one or both of the following: not immediately contiguous with both of the coding sequences with which it is immediately contiguous (i.e., one at the 5' end and one at the 3' end) in the naturally-occurring genome of the organism from which the nucleic acid is derived; or which is substantially free of a nucleic acid with which it occurs in the organism from which the nucleic acid is derived. The term includes, for example, a recombinant DNA which is incorporated into a vector, e.g., into an autonomously replicating plasmid or virus, or into the genomic DNA of a prokaryote or eukaryote, or which exists as a separate molecule (e.g., a cDNA or a genomic DNA fragment produced by PCR or restriction endonuclease treatment) independent of other DNA sequences. Substantially pure DNA also includes a recombinant DNA which is part of a hybrid gene encoding additional *H. pylori* DNA sequence.

A "contig" as used herein is a nucleic acid representing a continuous stretch of genomic sequence of an organism.

An "open reading frame", also referred to herein as ORF, is a region of nucleic acid which encodes a polypeptide. This region may represent a portion of a coding sequence or a total sequence.

As used herein, a "coding sequence" is a nucleic acid which is transcribed into messenger RNA and/or translated into a polypeptide when placed under the control of appropriate regulatory sequences. The boundaries of the coding sequence are determined by a translation start codon at the five prime terminus and a translation stop code at the three prime terminus. A coding sequence can include but is not limited to messenger RNA, synthetic DNA, and recombinant nucleic acid sequences.

A "complement" of a nucleic acid as used herein refers to an anti-parallel or antisense sequence that participates in Watson-Crick base-pairing with the original sequence.

A "gene product" is a protein or structural RNA which is specifically encoded for by a gene.

As used herein, the term "probe" refers to a nucleic acid, peptide or other chemical entity which specifically binds to a molecule of interest. Probes are often associated with or capable of associating with a label. A label is a chemical moiety capable of detection. Typical labels comprise dyes, radioisotopes, luminescent and chemiluminescent moieties, fluorophores, enzymes, precipitating agents, amplification sequences, and the like. Similarly, a nucleic acid, peptide or other chemical entity which specifically binds to a molecule of interest and immobilizes such molecule is referred herein as a "capture ligand". Capture ligands are typically associated with or capable of associating with a support such as nitro-cellulose, glass, nylon membranes, beads, particles and the like. The specificity of hybridization is dependent on conditions such as the base pair composition of the nucleotides, and the temperature and salt concentration of the reaction. These conditions are readily discernable to one of ordinary skill in the art using routine experimentation.

Homologous refers to the sequence similarity or sequence identity between two polypeptides or between two nucleic acid molecules. When a position in both of the two compared sequences is occupied by the same base or amino acid monomer subunit, e.g., if a position in each of two DNA molecules is occupied by adenine, then the molecules are homologous at that position. The percent of homology between two sequences is a function of the number of matching or homologous positions shared by the two sequences divided by the number of positions compared x 100. For example, if 6 of 10 of the positions in two sequences are matched or homologous then the two sequences are 60% homologous. By way of example, the DNA sequences ATTGCC and TATGGC

- 29 -

share 50% homology. Generally, a comparison is made when two sequences are aligned to give maximum homology.

Nucleic acids are hybridizable to each other when at least one strand of a nucleic acid can anneal to the other nucleic acid under defined stringency conditions.

- 5 Stringency of hybridization is determined by: (a) the temperature at which hybridization and/or washing is performed; and (b) the ionic strength and polarity of the hybridization and washing solutions. Hybridization requires that the two nucleic acids contain complementary sequences; depending on the stringency of hybridization, however, mismatches may be tolerated. Typically, hybridization of two sequences at high
- 10 stringency (such as, for example, in a solution of 0.5X SSC, at 65° C) requires that the sequences be essentially completely homologous. Conditions of intermediate stringency (such as, for example, 2X SSC at 65 ° C) and low stringency (such as, for example 2X SSC at 55° C), require correspondingly less overall complementarity between the hybridizing sequences. (1X SSC is 0.15 M NaCl, 0.015 M Na citrate).

- 15 The terms peptides, proteins, and polypeptides are used interchangeably herein.

As used herein, the term "surface protein" refers to all surface accessible proteins, e.g. inner and outer membrane proteins, proteins adhering to the cell wall, and secreted proteins.

- A polypeptide has *H. pylori* biological activity if it has one, two and preferably
- 20 more of the following properties: (1) if when expressed in the course of an *H. pylori* infection, it can promote, or mediate the attachment of *H. pylori* to a cell; (2) it has an enzymatic activity characteristic of an *H. pylori* protein; (3) or the gene which encodes it can rescue a lethal mutation in an *H. pylori* gene. A polypeptide has biological activity if it is an antagonist, agonist, or super-agonist of a polypeptide having one of the above-
- 25 listed properties.

- A biologically active fragment or analog is one having an *in vivo* or *in vitro* activity which is characteristic of the *H. pylori* polypeptides shown in the Sequence Listing, or of other naturally occurring *H. pylori* polypeptides, e.g., one or more of the biological activities described herein. Especially preferred are fragments which exist *in*
- 30 *vivo*, e.g., fragments which arise from post transcriptional processing or which arise from translation of alternatively spliced RNA's. Fragments include those expressed in native or endogenous cells as well as those made in expression systems, e.g., in CHO cells. Because peptides such as *H. pylori* polypeptides often exhibit a range of physiological properties and because such properties may be attributable to different
- 35 portions of the molecule, a useful *H. pylori* fragment or *H. pylori* analog is one which exhibits a biological activity in any biological assay for *H. pylori* activity. Most

- 30 -

preferably the fragment or analog possesses 10%, preferably 40%, more preferably 90% or greater of the activity of *H. pylori*, in any *in vivo* or *in vitro* assay.

5       Analogues can differ from naturally occurring *H. pylori* polypeptides in amino acid sequence or in ways that do not involve sequence, or both. Non-sequence modifications include changes in acetylation, methylation, phosphorylation, carboxylation, or glycosylation. Preferred analogues include *H. pylori* polypeptides (or biologically active fragments thereof) whose sequences differ from the wild-type sequence by one or more conservative amino acid substitutions or by one or more non-conservative amino acid substitutions, deletions, or insertions which do not substantially diminish the biological activity of the *H. pylori* polypeptide. Conservative substitutions typically include the substitution of one amino acid for another with similar characteristics, e.g., substitutions within the following groups: valine, glycine; glycine, alanine; valine, isoleucine, leucine; aspartic acid, glutamic acid; asparagine, glutamine; serine, threonine; lysine, arginine; and phenylalanine, tyrosine. Other conservative substitutions are outlined in 15       Table 2 below.

**TABLE 2**  
**CONSERVATIVE AMINO ACID REPLACEMENTS**

For Amino Acid	Code	Replace with any of
Alanine	A	D-Ala, Gly, beta-Ala, L-Cys, D-Cys
Arginine	R	D-Arg, Lys, D-Lys, homo-Arg, D-homo-Arg, Met, Ile, D-Met, D-Ile, Orn, D-Orn
Asparagine	N	D-Asn, Asp, D-Asp, Glu, D-Glu, Gln, D-Gln
Aspartic Acid	D	D-Asp, D-Asn, Asn, Glu, D-Glu, Gln, D-Gln
Cysteine	C	D-Cys, S-Me-Cys, Met, D-Met, Thr, D-Thr
Glutamine	Q	D-Gln, Asn, D-Asn, Glu, D-Glu, Asp, D-Asp
Glutamic Acid	E	D-Glu, D-Asp, Asp, Asn, D-Asn, Gln, D-Gln
Glycine	G	Ala, D-Ala, Pro, D-Pro, $\beta$ -Ala, Acp
Isoleucine	I	D-Ile, Val, D-Val, Leu, D-Leu, Met, D-Met
Leucine	L	D-Leu, Val, D-Val, Leu, D-Leu, Met, D-Met
Lysine	K	D-Lys, Arg, D-Arg, homo-Arg, D-homo-Arg, Met, D-Met, Ile, D-Ile, Orn, D-Orn
Methionine	M	D-Met, S-Me-Cys, Ile, D-Ile, Leu, D-Leu, Val, D-Val
Phenylalanine	F	D-Phe, Tyr, D-Thr, L-Dopa, His, D-His, Trp, D-Trp, Trans-3,4, or 5-phenylproline, cis-3,4, or 5-phenylproline
Proline	P	D-Pro, L-1-thioazolidine-4-carboxylic acid, D-or L-1-oxazolidine-4-carboxylic acid
Serine	S	D-Ser, Thr, D-Thr, allo-Thr, Met, D-Met, Met(O), D-Met(O), L-Cys, D-Cys
Threonine	T	D-Thr, Ser, D-Ser, allo-Thr, Met, D-Met, Met(O), D-Met(O), Val, D-Val
Tyrosine	Y	D-Tyr, Phe, D-Phe, L-Dopa, His, D-His
Valine	V	D-Val, Leu, D-Leu, Ile, D-Ile, Met, D-Met

Other analogs within the invention are those with modifications which increase peptide stability; such analogs may contain, for example, one or more non-peptide bonds (which replace the peptide bonds) in the peptide sequence. Also included are: analogs that include residues other than naturally occurring L-amino acids, e.g., D-amino acids  
5 or non-naturally occurring or synthetic amino acids, e.g.,  $\beta$  or  $\gamma$  amino acids; and cyclic analogs.

As used herein, the term "fragment", as applied to an *H. pylori* analog, will ordinarily be at least about 20 residues, more typically at least about 40 residues, preferably at least about 60 residues in length. Fragments of *H. pylori* polypeptides can  
10 be generated by methods known to those skilled in the art. The ability of a candidate fragment to exhibit a biological activity of *H. pylori* polypeptide can be assessed by methods known to those skilled in the art as described herein. Also included are *H. pylori* polypeptides containing residues that are not required for biological activity of the peptide or that result from alternative mRNA splicing or alternative protein processing  
15 events.

An "immunogenic component" as used herein is a moiety, such as an *H. pylori* polypeptide, analog or fragment thereof, that is capable of eliciting a humoral and/or cellular immune response in a host animal.

An "antigenic component" as used herein is a moiety, such as an *H. pylori*  
20 polypeptide, analog or fragment thereof, that is capable of binding to a specific antibody with sufficiently high affinity to form a detectable antigen-antibody complex.

As used herein, the term "transgene" means a nucleic acid (encoding, e.g., one or more polypeptides), which is partly or entirely heterologous, i.e., foreign, to the transgenic animal or cell into which it is introduced, or, is homologous to an endogenous  
25 gene of the transgenic animal or cell into which it is introduced, but which is designed to be inserted, or is inserted, into the cell's genome in such a way as to alter the genome of the cell into which it is inserted (e.g., it is inserted at a location which differs from that of the natural gene or its insertion results in a knockout). A transgene can include one or more transcriptional regulatory sequences and any other nucleic acid, such as introns,  
30 that may be necessary for optimal expression of the selected nucleic acid, all operably linked to the selected nucleic acid, and may include an enhancer sequence.

As used herein, the term "transgenic cell" refers to a cell containing a transgene.

As used herein, a "transgenic animal" is any animal in which one or more, and preferably essentially all, of the cells of the animal includes a transgene. The transgene  
35 can be introduced into the cell, directly or indirectly by introduction into a precursor of the cell, by way of deliberate genetic manipulation, such as by microinjection or by

- 32 -

infection with a recombinant virus. This molecule may be integrated within a chromosome, or it may be extrachromosomally replicating DNA.

The term "antibody" as used herein is intended to include fragments thereof which are specifically reactive with *H. pylori* polypeptides.

5 As used herein, the term "cell-specific promoter" means a DNA sequence that serves as a promoter, i.e., regulates expression of a selected DNA sequence operably linked to the promoter, and which effects expression of the selected DNA sequence in specific cells of a tissue. The term also covers so-called "leaky" promoters, which regulate expression of a selected DNA primarily in one tissue, but cause expression in  
10 other tissues as well.

Misexpression, as used herein, refers to a non-wild type pattern of gene expression. It includes: expression at non-wild type levels, i.e., over or under expression; a pattern of expression that differs from wild type in terms of the time or stage at which the gene is expressed, e.g., increased or decreased expression (as  
15 compared with wild type) at a predetermined developmental period or stage; a pattern of expression that differs from wild type in terms of decreased expression (as compared with wild type) in a predetermined cell type or tissue type; a pattern of expression that differs from wild type in terms of the splicing size, amino acid sequence, post-translational modification, or biological activity of the expressed polypeptide; a pattern of  
20 expression that differs from wild type in terms of the effect of an environmental stimulus or extracellular stimulus on expression of the gene, e.g., a pattern of increased or decreased expression (as compared with wild type) in the presence of an increase or decrease in the strength of the stimulus.

As used herein, "host cells" and other such terms denoting microorganisms or  
25 higher eukaryotic cell lines cultured as unicellular entities refers to cells which can become or have been used as recipients for a recombinant vector or other transfer DNA, and include the progeny of the original cell which has been transfected. It is understood by individuals skilled in the art that the progeny of a single parental cell may not necessarily be completely identical in genomic or total DNA compliment to the original  
30 parent, due to accident or deliberate mutation.

As used herein, the term "control sequence" refers to a nucleic acid having a base sequence which is recognized by the host organism to effect the expression of encoded sequences to which they are ligated. The nature of such control sequences differs depending upon the host organism; in prokaryotes, such control sequences generally  
35 include a promoter, ribosomal binding site and terminators; in eukaryotes, generally such control sequences include promoters, terminators and in some instances, enhancers.

- 33 -

The term control sequence is intended to include at a minimum, all components whose presence is necessary for expression, and may also include additional components whose presence is advantageous, for example, leader sequences.

As used herein, the term "operably linked" refers to sequences joined or ligated to function in their intended manner. For example, a control sequence is operably linked to coding sequence by ligation in such a way that expression of the coding sequence is achieved under conditions compatible with the control sequence and host cell.

The metabolism of a substance, as used herein, means any aspect of the, expression, function, action, or regulation of the substance. The metabolism of a substance includes modifications, e.g., covalent or non-covalent modifications of the substance. The metabolism of a substance includes modifications, e.g., covalent or non-covalent modification, the substance induces in other substances. The metabolism of a substance also includes changes in the distribution of the substance. The metabolism of a substance includes changes the substance induces in the distribution of other substances.

A "sample" as used herein refers to a biological sample, such as, for example, tissue or fluid isolated from an individual (including without limitation plasma, serum, cerebrospinal fluid, lymph, tears, saliva and tissue sections) or from *in vitro* cell culture constituents, as well as samples from the environment.

The practice of the invention will employ, unless otherwise indicated, conventional techniques of chemistry, molecular biology, microbiology, recombinant DNA, and immunology, which are within the skill of the art. Such techniques are explained fully in the literature. See e.g., Sambrook, Fritsch, and Maniatis, *Molecular Cloning; Laboratory Manual* 2nd ed. (1989); *DNA Cloning*, Volumes I and II (D.N Glover ed. 1985); *Oligonucleotide Synthesis* (M.J. Gait ed, 1984); *Nucleic Acid Hybridization* (B.D. Hames & S.J. Higgins eds. 1984); the series, *Methods in Enzymology* (Academic Press, Inc.), particularly Vol. 154 and Vol. 155 (Wu and Grossman, eds.) and *PCR-A Practical Approach* (McPherson, Quirke, and Taylor, eds., 1991).

#### I. Isolation of Nucleic Acids of *H. pylori* and Uses Therefor

##### *H. pylori* Genomic Sequence

This invention provides nucleotide sequences of the genome of *H. pylori* which thus comprises a DNA sequence library of *H. pylori* genomic DNA. The detailed description that follows provides nucleotide sequences of *H. pylori*, and also describes how the sequences were obtained and how ORFs and protein-coding sequences were

- 34 -

identified. Also described are methods of using the disclosed *H. pylori* sequences in methods including diagnostic and therapeutic applications. Furthermore, the library can be used as a database for identification and comparison of medically important sequences in this and other strains of *H. pylori*.

5 To determine the genomic sequence of *H. pylori*, DNA was isolated from a strain of *H. pylori* and mechanically sheared by nebulization to a median size of 2 kb. Following size fractionation by gel electrophoresis, the fragments were blunt-ended, ligated to adapter oligonucleotides, and cloned into each of 20 different pMPX vectors (Rice et al., abstracts of Meeting of Genome Mapping and Sequencing, Cold Spring  
10 Harbor, NY, 5/11-5/15, 1994, p. 225) to construct a series of "shotgun" subclone libraries.

DNA sequencing was achieved using multiplex sequencing procedures essentially as disclosed in Church et al., 1988, *Science* 240:185; U.S. Patents No. 4,942,124 and 5,149,625). DNA was extracted from pooled cultures and subjected to  
15 chemical or enzymatic sequencing. Sequencing reactions were resolved by electrophoresis, and the products were transferred and covalently bound to nylon membranes. Finally, the membranes were sequentially hybridized with a series of labelled oligonucleotides complimentary to "tag" sequences present in the different shotgun cloning vectors. In this manner, a large number of sequences could be obtained  
20 from a single set of sequencing reactions. The cloning and sequencing procedures are described in more detail in the Exemplification.

Individual sequence reads obtained in this manner were assembled using the FALCON™ program (Church et al., 1994, *Automated DNA Sequencing and Analysis*, J.C. Venter, ed., Academic Press) and PHRAP (P. Green, Abstracts of DOE Human  
25 Genome Program Contractor-Grantee Workshop V, Jan. 1996, p.157). A resulting assembly of contigs, each representing a continuous stretch of DNA or DNA sequence was obtained. The average contig length was about 3 kb.

A variety of approaches are used to order the contigs so as to obtain a continuous sequence representing the entire *H. pylori* genome. Synthetic oligonucleotides are  
30 designed that are complementary to sequences at the end of each contig. These oligonucleotides may be hybridized to libraries of *H. pylori* genomic DNA in, for example, lambda phage vectors or plasmid vectors to identify clones that contain sequences corresponding to the junctional regions between individual contigs. Such clones are then used to isolate template DNA and the same oligonucleotides are used as  
35 primers in polymerase chain reaction (PCR) to amplify junctional fragments, the nucleotide sequence of which was then determined.

- 35 -

The *H. pylori* sequences were analyzed for the presence of open reading frames (ORFs) comprising at least 180 nucleotides. ORFs of at least 180 nucleotides (based on stop-to-stop codon reads) were predicted. As a result of the analysis of ORFs based on stop-to-stop codon reads, it should be understood that these ORFs may not correspond to the ORF of a naturally-occurring *H. pylori* polypeptide. These ORFs may contain start codons which indicate the initiation of protein synthesis of a naturally-occurring *H. pylori* polypeptide. Such start codons within the ORFs provided herein can be identified by those of ordinary skill in the relevant art and the resulting ORF and the encoded *H. pylori* polypeptide is within the scope of this invention. For example, within the ORFs a codon such as AUG or GUG (encoding methionine or valine) which is part of the initiation signal for protein synthesis can be identified and the ORF modified to correspond to a naturally-occurring *H. pylori* polypeptide. The predicted coding regions were defined by evaluating the coding potential of such sequences with the program GENEMARK™ (Borodovsky and McIninch, 1993, *Comp. Chem.* 17:123).

#### Other *H. pylori* Nucleic Acids

The nucleic acids of this invention may be obtained directly from the DNA of the above referenced *H. pylori* strain by using the polymerase chain reaction (PCR). See "PCR, A Practical Approach" (McPherson, Quirke, and Taylor, eds., IRL Press, Oxford, UK, 1991) for details about the PCR. High fidelity PCR can be used to ensure a faithful DNA copy prior to expression. In addition, amplified products can be checked by conventional sequencing methods. Clones carrying the desired sequences described in this invention may be obtained by screening the libraries by means of the PCR or by hybridization of synthetic oligonucleotide probes to filter lifts of the library colonies or plaques as known in the art (see, e.g., Sambrook et al., *Molecular Cloning, A Laboratory Manual* 2nd edition, 1989, Cold Spring Harbor Press, NY).

It is also possible to obtain nucleic acids encoding *H. pylori* polypeptides from a cDNA library in accordance with protocols herein described. A cDNA encoding an *H. pylori* polypeptide can be obtained by isolating total mRNA from an appropriate cell line. Double stranded cDNAs can then be prepared from the total mRNA.

Subsequently, the cDNAs can be inserted into a suitable plasmid or viral (e.g., bacteriophage) vector using any one of a number of known techniques. Genes encoding *H. pylori* polypeptides can also be cloned using established polymerase chain reaction techniques in accordance with the nucleotide sequence information provided by the invention. The nucleic acids of the invention can be DNA or RNA. Preferred nucleic acids are shown in the Sequence Listing.

- 36 -

The nucleic acids of the invention can also be chemically synthesized using standard techniques. Various methods of chemically synthesizing polydeoxynucleotides are known, including solid-phase synthesis which, like peptide synthesis, has been fully automated in commercially available DNA synthesizers (See e.g., Itakura et al. U.S. Patent No. 4,598,049; Caruthers et al. U.S. Patent No. 4,458,066; and Itakura U.S. Patent Nos. 4,401,796 and 4,373,071, incorporated by reference herein).

Nucleic acids isolated or synthesized in accordance with features of the present invention are useful, by way of example, without limitation, as probes, primers, capture ligands, antisense genes and for developing expression systems for the synthesis of proteins and peptides corresponding to such sequences. As probes, primers, capture ligands and antisense agents, the nucleic acid normally consists of all or part (approximately twenty or more nucleotides for specificity as well as the ability to form stable hybridization products) of the nucleic acids shown in the Sequence Listing. These uses are described in further detail below.

#### Probes

A nucleic acid isolated or synthesized in accordance with the nucleotide sequences set forth in the Sequence Listing can be used as a probe to specifically detect *H. pylori*. With the sequence information set forth in the present application, sequences of twenty or more nucleotides are identified which provide the desired inclusivity and exclusivity with respect to *H. pylori*, and extraneous nucleic acids likely to be encountered during hybridization conditions. More preferably, the sequence will comprise at least twenty to thirty nucleotides to convey stability to the hybridization product formed between the probe and the intended target molecules.

Sequences larger than 1000 nucleotides in length are difficult to synthesize but can be generated by recombinant DNA techniques. Individuals skilled in the art will readily recognize that the nucleic acids, for use as probes, can be provided with a label to facilitate detection of a hybridization product.

Nucleic acid isolated and synthesized in accordance with the Sequence Listing can also be useful as probes to detect homologous regions (especially homologous genes) of other *Helicobacter* species using appropriate stringency hybridization conditions as described herein.

#### Capture Ligand

For use as a capture ligand, the nucleic acid selected in the manner described above with respect to probes, can be readily associated with a support. The manner in which nucleic acid is associated with supports is well known. Nucleic acid having twenty or more nucleotides in a sequence contained in the Sequence Listing have utility

- 37 -

to separate *H. pylori* nucleic acid from the nucleic acid of each other and other organisms. Nucleic acid having twenty or more nucleotides in a sequence shown in the Sequence Listing can also have utility to separate other *Helicobacter* species from each other and from other organisms. Preferably, the sequence will comprise at least twenty  
5 nucleotides to convey stability to the hybridization product formed between the probe and the intended target molecules. Sequences larger than 1000 nucleotides in length are difficult to synthesize but can be generated by recombinant DNA techniques.

#### Primers

Nucleic acid isolated or synthesized in accordance with the sequences described  
10 herein have utility as primers for the amplification of *H. pylori* nucleic acid. These nucleic acids may also have utility as primers for the amplification of nucleic acids in other *Helicobacter* species. With respect to polymerase chain reaction (PCR) techniques, nucleic acids of  $\geq 10$ -15 nucleotides contained in the Sequence Listing have utility in conjunction with suitable enzymes and reagents to create copies of *H. pylori*  
15 nucleic acid. More preferably, the sequence will comprise twenty or more nucleotides to convey stability to the hybridization product formed between the primer and the intended target molecules. Binding conditions of primers greater than 100 nucleotides are more difficult to control to obtain specificity. High fidelity PCR can be used to ensure a faithful DNA copy prior to expression. In addition, amplified products can be  
20 checked by conventional sequencing methods.

The copies can be used in diagnostic assays to detect specific sequences, including genes from *H. pylori* and/or other *Helicobacter* species. The copies can also be incorporated into cloning and expression vectors to generate polypeptides corresponding to the nucleic acid synthesized by PCR, as is described in greater detail  
25 herein.

#### Antisense

Nucleic acid or nucleic acid-hybridizing derivatives isolated or synthesized in accordance with the sequences described herein have utility as antisense agents to prevent the expression of *H. pylori* genes. These sequences also have utility as antisense  
30 agents to prevent expression of genes of other *Helicobacter* species.

In one embodiment, nucleic acid or derivatives corresponding to *H. pylori* nucleic acids is loaded into a suitable carrier such as a liposome or bacteriophage for introduction into bacterial cells. For example, a nucleic acid having twenty or more nucleotides is capable of binding to bacteria nucleic acid or bacteria messenger RNA.  
35 Preferably, the antisense nucleic acid is comprised of 20 or more nucleotides to provide necessary stability of a hybridization product of non-naturally occurring nucleic acid and

- 38 -

bacterial nucleic acid and/or bacterial messenger RNA. Nucleic acid having a sequence greater than 1000 nucleotides in length is difficult to synthesize but can be generated by recombinant DNA techniques. Methods for loading antisense nucleic acid in liposomes is known in the art as exemplified by U.S. Patent 4,241,046 issued December 23, 1980 to Papahadjopoulos et al.

## II. Expression of *H. pylori* Nucleic Acids

Nucleic acid isolated or synthesized in accordance with the sequences described herein have utility to generate polypeptides. The nucleic acids exemplified in the Sequence Listing or fragments of said nucleic acid encoding active portions of *H. pylori* polypeptides can be cloned into suitable vectors or used to isolate nucleic acid. The isolated nucleic acid is combined with suitable DNA linkers and cloned into a suitable vector.

The function of a specific gene or operon can be ascertained by expression in a bacterial strain under conditions where the activity of the gene product(s) specified by the gene or operon in question can be specifically measured. Alternatively, a gene product may be produced in large quantities in an expressing strain for use as an antigen, an industrial reagent, for structural studies, etc. This expression can be accomplished in a mutant strain which lacks the activity of the gene to be tested, or in a strain that does not produce the same gene product(s). This includes, but is not limited to other *Helicobacter* strains, and other bacterial strains such as *E. coli*, *Norcardia*, *Corynebacterium*, and *Streptomyces* species. In some cases the expression host will utilize the natural *Helicobacter* promoter whereas in others, it will be necessary to drive the gene with a promoter sequence derived from the expressing organism (e.g., an *E. coli* beta-galactosidase promoter for expression in *E. coli*).

To express a gene product using the natural *H. pylori* promoter, a procedure such as the following can be used. A restriction fragment containing the gene of interest, together with its associated natural promoter element and regulatory sequences (identified using the DNA sequence data) is cloned into an appropriate recombinant plasmid containing an origin of replication that functions in the host organism and an appropriate selectable marker. This can be accomplished by a number of procedures known to those skilled in the art. It is most preferably done by cutting the plasmid and the fragment to be cloned with the same restriction enzyme to produce compatible ends that can be ligated to join the two pieces together. The recombinant plasmid is introduced into the host organism by, for example, electroporation and cells containing the

- 39 -

recombinant plasmid are identified by selection for the marker on the plasmid. Expression of the desired gene product is detected using an assay specific for that gene product.

In the case of a gene that requires a different promoter, the body of the gene (coding sequence) is specifically excised and cloned into an appropriate expression plasmid. This subcloning can be done by several methods, but is most easily accomplished by PCR amplification of a specific fragment and ligation into an expression plasmid after treating the PCR product with a restriction enzyme or exonuclease to create suitable ends for cloning.

A suitable host cell for expression of a gene can be any procaryotic or eucaryotic cell. For example, an *H. pylori* polypeptide can be expressed in bacterial cells such as *E. coli*, insect cells (baculovirus), yeast, or mammalian cells such as Chinese hamster ovary cell (CHO). Other suitable host cells are known to those skilled in the art.

Expression in eucaryotic cells such as mammalian, yeast, or insect cells can lead to partial or complete glycosylation and/or formation of relevant inter- or intra-chain disulfide bonds of a recombinant peptide product. Examples of vectors for expression in yeast *S. cerevisiae* include pYepSec1 (Baldari et al., (1987) *Embo J.* 6:229-234), pMFa (Kurjan and Herskowitz, (1982) *Cell* 30:933-943), pJRY88 (Schultz et al., (1987) *Gene* 54:113-123), and pYES2 (Invitrogen Corporation, San Diego, CA). Baculovirus vectors available for expression of proteins in cultured insect cells (SF 9 cells) include the pAc series (Smith et al., (1983) *Mol. Cell Biol.* 3:2156-2165) and the pVL series (Lucklow, V.A., and Summers, M.D., (1989) *Virology* 170:31-39). Generally, COS cells (Gluzman, Y., (1981) *Cell* 23:175-182) are used in conjunction with such vectors as pCDM 8 (Aruffo, A. and Seed, B., (1987) *Proc. Natl. Acad. Sci. USA* 84:8573-8577) for transient amplification/expression in mammalian cells, while CHO (dhfr<sup>-</sup> Chinese Hamster Ovary) cells are used with vectors such as pMT2PC (Kaufman et al. (1987), *EMBO J.* 6:187-195) for stable amplification/expression in mammalian cells. Vector DNA can be introduced into mammalian cells via conventional techniques such as calcium phosphate or calcium chloride co-precipitation, DEAE-dextran-mediated transfection, or electroporation. Suitable methods for transforming host cells can be found in Sambrook et al. (*Molecular Cloning: A Laboratory Manual*, 2nd Edition, Cold Spring Harbor Laboratory press (1989)), and other laboratory textbooks.

Expression in procaryotes is most often carried out in *E. coli* with either fusion or non-fusion inducible expression vectors. Fusion vectors usually add a number of NH<sub>2</sub> terminal amino acids to the expressed target gene. These NH<sub>2</sub> terminal amino acids often are referred to as a reporter group. Such reporter groups usually serve two purposes: 1) to increase the solubility of the target recombinant protein; and 2) to aid in

- 40 -

the purification of the target recombinant protein by acting as a ligand in affinity purification. Often, in fusion expression vectors, a proteolytic cleavage site is introduced at the junction of the reporter group and the target recombinant protein to enable separation of the target recombinant protein from the reporter group subsequent to purification of the fusion protein. Such enzymes, and their cognate recognition sequences, include Factor Xa, thrombin and enterokinase. Typical fusion expression vectors include pGEX (Amrad Corp., Melbourne, Australia), pMAL (New England Biolabs, Beverly, MA) and pRIT5 (Pharmacia, Piscataway, NJ) which fuse glutathione S-transferase, maltose E binding protein, or protein A, respectively, to the target recombinant protein. A preferred reporter group is poly(His), which may be fused to the amino or carboxy terminus of the protein and which renders the recombinant fusion protein easily purifiable by metal chelate chromatography.

Inducible non-fusion expression vectors include pTrc (Amann et al., (1988) *Gene* 69:301-315) and pET11d (Studier et al., Gene Expression Technology: Methods in Enzymology 185, Academic Press, San Diego, California (1990) 60-89). While target gene expression relies on host RNA polymerase transcription from the hybrid trp-lac fusion promoter in pTrc, expression of target genes inserted into pET11d relies on transcription from the T7 gn10-lac 0 fusion promoter mediated by coexpressed viral RNA polymerase (T7 gn1). This viral polymerase is supplied by host strains BL21(DE3) or HMS174(DE3) from a resident  $\lambda$  prophage harboring a T7 gn1 under the transcriptional control of the lacUV 5 promoter.

For example, a host cell transfected with a nucleic acid vector directing expression of a nucleotide sequence encoding an *H. pylori* polypeptide can be cultured under appropriate conditions to allow expression of the polypeptide to occur. The polypeptide may be secreted and isolated from a mixture of cells and medium containing the peptide. Alternatively, the polypeptide may be retained cytoplasmically and the cells harvested, lysed and the protein isolated. A cell culture includes host cells, media and other byproducts. Suitable media for cell culture are well known in the art. Polypeptides of the invention can be isolated from cell culture medium, host cells, or both using techniques known in the art for purifying proteins including ion-exchange chromatography, gel filtration chromatography, ultrafiltration, electrophoresis, and immunoaffinity purification with antibodies specific for such polypeptides. Additionally, in many situations, polypeptides can be produced by chemical cleavage of a native protein (e.g., tryptic digestion) and the cleavage products can then be purified by standard techniques.

- 41 -

In the case of membrane bound proteins, these can be isolated from a host cell by contacting a membrane-associated protein fraction with a detergent forming a solubilized complex, where the membrane-associated protein is no longer entirely embedded in the membrane fraction and is solubilized at least to an extent which allows it to be chromatographically isolated from the membrane fraction. Several different criteria are used for choosing a detergent suitable for solubilizing these complex. For example, one property considered is the ability of the detergent to solubilize the *H. pylori* protein within the membrane fraction at minimal denaturation of the membrane-associated protein allowing for the activity or functionality of the membrane-associated protein to return upon reconstitution of the protein. Another property considered when selecting the detergent is the critical micells concentration (CMC) of the detergent in that the detergent of choice preferably has a high CMC value allowing for ease of removal after reconstitution. A third property considered when selecting a detergent is the hydrophobicity of the detergent. Typically, membrane-associated proteins are very hydrophobic and therefore detergents which are also hydrophobic, e.g., the triton series, would be useful for solubilizing the hydrophobic proteins. Another property important to a detergent can be the capability of the detergent to remove the *H. pylori* protein with minimal protein-protein interaction facilitating further purification. A fifth property of the detergent which should be considered is the charge of the detergent. For example, if it is desired to use ion exchange resins in the purification process then preferably detergent should be an uncharged detergent. Chromatographic techniques which can be used in the final purification step are known in the art and include hydrophobic interaction, lectin affinity, ion exchange, dye affinity and immunoaffinity.

One strategy to maximize recombinant *H. pylori* peptide expression in *E. coli* is to express the protein in a host bacteria with an impaired capacity to proteolytically cleave the recombinant protein (Gottesman, S., Gene Expression Technology: Methods in Enzymology 185, Academic Press, San Diego, California (1990) 119-128). Another strategy would be to alter the nucleic acid encoding *H. pylori* peptide to be inserted into an expression vector so that the individual codons for each amino acid would be those preferentially utilized in highly expressed *E. coli* proteins (Wada et al., (1992) *Nuc. Acids Res.* 20:2111-2118). Such alteration of nucleic acids of the invention can be carried out by standard DNA synthesis techniques.

The nucleic acids of the invention can also be chemically synthesized using standard techniques. Various methods of chemically synthesizing polydeoxynucleotides are known, including solid-phase synthesis which, like peptide synthesis, has been fully automated in commercially available DNA synthesizers (See, e.g., Itakura et al. U.S.

Patent No. 4,598,049; Caruthers et al. U.S. Patent No. 4,458,066; and Itakura U.S. Patent Nos. 4,401,796 and 4,373,071, incorporated by reference herein).

### III. *H. pylori* Polypeptides

5 This invention encompasses isolated *H. pylori* polypeptides encoded by the disclosed *H. pylori* genomic sequences, including the polypeptides contained in the Sequence Listing. Polypeptides of the invention are preferably at least 5 amino acid residues in length. Using the DNA sequence information provided herein, the amino acid sequences of the polypeptides encompassed by the invention can be deduced using  
10 methods well-known in the art. It will be understood that the sequence of an entire nucleic acid encoding an *H. pylori* polypeptide can be isolated and identified based on an ORF that encodes only a fragment of the cognate protein-coding region. This can be achieved, for example, by using the isolated nucleic acid encoding the ORF, or fragments thereof, to prime a polymerase chain reaction with genomic *H. pylori* DNA as  
15 template; this is followed by sequencing the amplified product.

The polypeptides of the invention can be isolated from wild-type or mutant *H. pylori* cells or from heterologous organisms or cells (including, but not limited to, bacteria, yeast, insect, plant and mammalian cells) into which an *H. pylori* nucleic acid has been introduced and expressed. In addition, the polypeptides can be part of  
20 recombinant fusion proteins.

*H. pylori* polypeptides of the invention can be chemically synthesized using commercially automated procedures such as those referenced herein.

Many of the polypeptides of the invention are related to one another. Some of these relationships are described in Tables 3-6 below. All of the polypeptide lengths in  
25 Table 3 are from stop codon to stop codon in the nucleotide sequence of *H. pylori*. As is known in the art, the actual polypeptide lengths are usually shorter than the stop-to-stop codon lengths because a start codon for an initiator charged tRNA usually appears a few nucleotides downstream from the prior stop codon and within a few nucleotides following a ribosome binding site (also known as a "Shine-Delgarno sequence"). Since  
30 most of the ribosome binding sites in *H. pylori* have many of the same general features of those known in *E. coli*, one skilled in the art can predict the actual start codon with good reliability from the stop-to-stop nucleotide sequence of an open reading frame. The polypeptide sequences of SEQ ID NOs:492-743 of this invention represent the stop-to-stop codon lengths of the open reading frames of SEQ ID NOs:1-252. All other  
35 polypeptide sequences of this invention represent the predicted start to stop protein lengths from the nucleotide sequences. One skilled in the art can recognize start sites in

- 43 -

the stop-to-stop open reading frames of the nucleotide sequences presented herein. In addition, one skilled in the art will occasionally detect alternative start sites, some of which may be utilized *in vivo* by the cellular machinery. The number of these alternative start sites is sufficiently small that one skilled in the art can readily test them in a recombinant expression systems known in the art to determine which ones provide authentic functional protein products.

The relationship between the polypeptides shown in Table 3 can be described as follows. First, all of the polypeptides of Table 3 are at least 90% identical with each other over most of their lengths, and most are over 95% identical with each other. Second, the stop-to-stop lengths are different for some of the homologous pairs of polypeptides. In some cases, the shorter polypeptide contains the relevant portion of the protein exhibiting utility in this invention; in some cases, the longer polypeptide may exhibit improved utility. Third, some polypeptides in the second column are homologous to two shorter polypeptides in the fifth column.

In all cases, the homology relationships described in Table 3 are highly significant. For example, a typical *H. pylori* gene product will exhibit amino acid sequence identities of between 92% and 100% among different strains of *H. pylori* selected from human patients. The nucleotide sequences encoding the related polypeptides of this invention are also very similar to one another. For example, nucleotide probes derived from the coding sequence of a polypeptide of this invention can be used in PCR or hybridization experiments to identify clones carrying the nucleotide sequence encoding the homologous related polypeptide.

-44-

TABLE 3

ORF Name	aa Seq ID #	Length	ORF Name	aa Seq ID #	Length	Length Difference	%ID	Overlap
01ae11010_40688_c2_38	745	261	hp3e11075orf3	632	261	0	100%	261
01ae12001_24218781_f2_18	746	148	05gp11901orf19	553	370	-222	95%	149
01ae12001_24218781_f2_18	746	148	07gp11807orf17	568	195	-47	100%	143
01ge10203_35281542_c3_16	797	261	01ge10203orf7	659	154	107	91%	138
01ge10203_35281542_c3_16	797	261	01ge10203orf14	657	71	190	99%	69
01ge10203_860166_f3_9	798	233	01ge10203orf6	658	233	0	100%	233
01ge11619_13788141_c2_11	799	105	11ge10309orf4	687	105	0	100%	105
01ge11619_23711062_c3_14	747	319	11ge10309orf12	579	319	0	100%	319
01ge11619_24415880_c2_12	800	173	11ge10309orf5	688	173	0	100%	173
01ge11619_24417813_c1_8	801	324	11ge10309orf24	684	324	0	100%	324
01gp11016_4103403_c2_13	748	145	01gp11016orf13	508	123	22	100%	123
02ae11612_1074212_f1_1	749	103	02ae11612orf1	509	103	0	100%	103
02ae11612_22477267_f2_27	750	304	03xe11215orf7	529	152	152	94%	148
02ae11612_23598175_f1_2	751	332	02ae11612orf15	510	332	0	100%	332
02ae11612_33203250_c1_51	752	509	05cp11911orf35	549	206	303	99%	205
02cp20506_24611325_f2_6	835	256	01ce11513orf17	500	256	0	100%	256
02ge10116_36367936_c1_19	753	173	02ge20116orf25	521	173	0	100%	173
02ge11622_21693936_c1_54	839	313	13ee10216orf57	743	313	0	98%	313
02ge11622_875260_f3_36	754	523	13ee10216orf7	586	523	0	100%	523
02gp20706_1203402_c3_58	755	162	09ge70821orf2	574	162	0	100%	162
02gp20706_15781452_c2_51	756	97	06ge10115orf17	563	97	0	100%	97
02gp20706_16803513_f1_1	802	488	01cp11710orf16	652	64	424	98%	63
02gp20706_16803513_f1_1	802	488	01cp11710orf6	655	247	241	99%	247
02gp20706_16803513_f1_1	802	488	01cp11710orf5	654	146	342	99%	145
02gp20706_20365905_f2_8	803	198	01cp11710orf9	656	130	68	95%	130
02gp20706_20365905_f2_8	803	198	01cp11710orf18	653	85	113	100%	75
02gp20706_23632775_f3_32	757	230	06ge10115orf4	565	230	0	100%	230

TABLE 3 (continued)

02gp20706_23866562_c2_53	758	120	06ge10115orf19	564	84	36	96%	84
02gp20706_4892558_f3_19	759	218	01cp11710orf1	502	218	0	100%	218
02gp20814_24415958_f3_9	760	176	14ee10308orf4	594	176	0	99%	176
02gp20814_3984818_f1_1	804	215	14ee10308orf1	593	68	147	97%	67
02gp20814_3984818_f1_1	804	215	14ee10308orf7	694	155	60	100%	143
04cp11202_16603425_c2_72	805	185	14cp10119orf1	692	161	24	96%	158
04cp11202_19797128_f1_5	806	199	04ge11713orf10	665	132	67	100%	127
04cp11202_23553177_c1_75	807	372	06ap10209orf4	674	88	284	96%	79
04cp11202_23553177_c1_75	807	372	06ap10209orf1	673	186	186	98%	179
04cp11202_23553177_c3_109	808	372	06ap10209orf4	674	88	284	96%	79
04cp11202_23553177_c3_109	808	372	06ap10209orf1	673	186	186	98%	179
04cp11202_24256567_c3_117	744	536	04ge11713orf35	601	200	336	99%	195
04cp11202_24256567_c3_117	744	536	04ge11713orf28	666	122	414	91%	111
04cp11202_24256567_c3_117	744	536	04ge11713orf36	667	250	286	97%	234
04cp11202_24261588_f2_23	761	96	01ae12021orf5	492	96	0	100%	96
04ge10816_22086531_f2_10	809	208	04ge11210orf1	664	208	0	100%	208
04ge10816_33726080_c2_29	810	369	13ae10610orf1	690	133	236	98%	120
05ae30220_21619067_f3_56	763	293	05ae20220orf58	544	293	0	100%	293
05ae30220_24882812_c3_103	765	486	05ae20220orf119	542	486	0	100%	486
05ae30220_25953163_c3_98	766	81	05ae20220orf95	545	81	0	100%	81
05ae30220_9882767_f2_34	812	69	05ae20220orf37	671	69	0	100%	69
05cp21223_4725443_f3_14	813	113	hpl10506orf5	696	113	0	100%	113
05ee10816_14649077_f3_18	767	263	04ce11617orf4	534	152	111	97%	144
05ee10816_259703_f2_7	814	158	04ce11617orf10	661	142	16	93%	142
05ee10816_4103408_f2_11	768	115	04ce11617orf16	531	115	0	100%	115
05ee10816_4687651_c1_22	769	953	04ce11617orf27	519	676	277	100%	676
05ee10816_4687651_c1_22	769	953	04ce11617orf26	532	238	715	100%	238
06ap11119_16594193_f1_9	770	87	11ge10309orf7	580	87	0	100%	87
06ap11119_24426508_f3_26	815	120	11ge10309orf28	685	137	-17	100%	121
06ap11119_24426508_f3_26	815	120	hp4e12063orf1	699	86	34	100%	65
06ap20306_23437632_f3_9	771	291	04gp11803orf13	541	243	48	99%	228

-46-

TABLE 3 (continued)

06cp30603_23452_c3_80	772	176	560	159	17	97%	155
06cp30603_23476568_c1_44	773	442	550	442	0	100%	442
06cp30603_4689068_c3_79	774	100	559	100	0	100%	100
06ee10709_21675012_f1_2	816	284	675	284	0	100%	284
07ap11015_23938312_f3_2	762	271	622	95	176	100%	94
07ap11015_23938312_f3_2	762	271	624	211	60	99%	177
07ap1111_234693_c3_14	775	209	566	209	0	100%	209
07ap11213_35156577_c1_24	836	401	718	420	-19	99%	399
07ap11213_35401528_c1_21	764	803	672	458	345	100%	458
07ap11213_35401528_c1_21	764	803	620	344	459	100%	344
07ap20216_7227202_f3_10	776	532	547	529	3	100%	529
07ap80601_976413_f3_9	837	274	633	268	6	100%	265
07ce11019_22051291_f1_1	817	600	678	193	407	99%	190
07ce11019_22051291_f1_1	817	600	679	418	182	100%	414
07ep11916_5273452_c3_31	777	255	501	255	0	100%	255
07ep11916_5913592_f3_18	818	272	654	146	126	93%	111
09cp11003_19532625_c3_17	778	357	599	357	0	99%	357
09cp11003_5945252_f2_4	819	219	695	219	0	100%	219
09cp20502_24001388_c1_31	779	61	600	61	0	100%	61
09ze10333_22460750_f2_6	820	729	660	323	406	97%	312
12ge10321_24308513_f3_20	840	242	741	242	0	99%	242
12ge10321_4821082_f3_14	821	172	683	115	57	100%	115
12ge10321_4821082_f3_14	821	172	701	68	104	92%	64
12gp31106_3024126_f2_25	780	147	585	147	0	100%	147
13ae10712_14100018_f2_12	781	201	592	201	0	99%	201
13ae10712_29569208_c2_27	782	264	584	264	0	100%	264
14ap10221_13689381_c3_4	822	366	553	370	-4	96%	360
14ap10221_13689381_c3_4	822	366	676	66	300	100%	66
14ap10815_20585777_c1_13	783	677	628	677	0	100%	677
14ce21516_85786_f1_1	841	700	533	361	339	98%	350
14ce21516_85786_f1_1	841	700	740	284	416	95%	271

-47-

TABLE 3 (continued)

29ep10720_24220926_f2_8	823	86	hp4e12063orf1	699	86	0	100%	86
29ep10720_24220926_f2_8	823	86	11ge10309orf28	685	137	-51	100%	81
29ep10720_24432762_c3_39	824	227	11ge10309orf39	686	227	0	100%	227
29ge30321_12913562_f1_1	825	79	01ce11618orf22	648	79	0	100%	79
29ge30321_135253_f2_6	826	101	01ce11618orf7	650	101	0	97%	101
29ge30321_21673965_f2_7	827	488	01ce11618orf9	651	409	79	100%	409
29ge30321_24336712_f1_5	828	422	01ce11618orf11	647	188	234	93%	181
29ge30321_24336712_f1_5	828	422	01ce11618orf27	649	235	187	96%	235
29ge30321_34157812_f3_10	784	342	01ce11618orf24	498	86	256	97%	85
hp1p14013_11726503_c2_20	785	149	hp1p14013orf17	617	149	0	99%	149
hp2p10272_22692325_f3_21	829	188	09ap20802orf13	680	163	25	99%	149
hp2p10272_23697200_f3_22	786	329	09ap20802orf14	570	278	51	100%	276
hp2p10272_24406280_c1_26	830	194	09ap20802orf21	681	194	0	100%	194
hp2p10272_26829136_f1_1	787	343	hp2p10272orf1	625	343	0	100%	343
hp3p10807_189075_f2_4	838	312	hp3p10807orf6	728	325	-13	99%	312
hp3p10807_29343768_f1_1	831	187	hp3p10807orf4	697	187	0	100%	187
hp3p10807_29352212_f2_5	832	230	hp3p10807orf7	698	230	0	100%	230
hp5e15211_819455_c2_24	788	269	hp5e15211orf23	641	269	0	100%	269
hp5p15212_34064750_f2_9	789	148	02ce10213orf19	515	148	0	98%	148
hp5p15641_21698387_c2_20	790	185	hp5p15641orf23	644	185	0	98%	185
hp6e10967_23476502_f2_6	791	165	hp2p10625orf5	626	165	0	99%	165
hp6e10967_24882750_f2_7	792	157	hp2p10625orf6	627	157	0	100%	157
hp6e12267_14650278_f3_29	842	277	12ge20305orf26	742	277	0	98%	277
hp6e12267_4876718_f2_23	793	143	12ge10305orf5	582	146	-3	98%	146
hp6e20339_1190660_c2_46	794	69	04gp11213orf22	540	69	0	100%	69
hp6e20339_21492187_c1_40	795	106	04gp11213orf14	539	136	-30	99%	106
hp6e20339_24317062_c3_57	833	382	04gp11213orf5	669	382	0	99%	382
hp6e20339_34024187_c1_37	796	122	04gp11213orf11	538	122	0	100%	122
hp6p10233_12273302_f1_1	834	143	07cp10312orf5	677	138	5	95%	105

- 48 -

Additional relationships between polypeptides of the invention are described in Table 4 below. All of the polypeptide lengths in Table 4 below are measured from stop codon to stop codon in the nucleotide sequence of *H. pylori*.

The relationship between the polypeptides shown in Table 4 can be described as follows. First, all of the polypeptides of Table 4 are at least 90% identical with each other over most of their lengths, and most are over 95% identical with each other. Second, the stop-to-stop lengths are different for some of the homologous pairs of polypeptides. In some cases, the shorter polypeptide contains the relevant portion of the protein exhibiting utility in this invention; in some cases, the longer polypeptide may exhibit improved utility. Third, some polypeptides in the second column are homologous to two shorter polypeptides in the fifth column.

In all cases, the homology relationships described in Table 4 are highly significant. For example, a typical *H. pylori* gene product will exhibit amino acid sequence identities of between 92% and 100% among different strains of *H. pylori* selected from human patients. The nucleotide sequences encoding the related polypeptides of this invention are also very similar to one another. For example, nucleotide probes derived from the coding sequence of a polypeptide of this invention can be used in PCR or hybridization experiments to identify clones carrying the nucleotide sequence encoding the homologous related polypeptide.

TABLE 4

ORF Name	aa Seq ID#	Length (aa)	ORF Name	aa Seq ID#	Length (aa)	% Iden- tity	Over- lap (aa)	Change in Length
01cel1104_36125337_c1_8	849	74	hp1e10523orf3	612	63	100	63	11
01ce21104_33203250_c3_87	850	509	05cp11911orf35	549	206	100	204	303
01cp20708_36134808_f2_11	928	230	14ee10419orf5	704	229	100	229	1
02ae31010_12504512_f3_28.aa	929	490	01ge10801orf2	514	60	95	56	430
02ae31010_16833312_f2_19	930	181	02cel1022orf2	530	184	98	181	-3
02ae31010_2117087_f3_34	931	137	07ce10203orf17	635	76	97	77	61
02ae31010_30208317_f1_14	932	343	hp1p13947orf10	717	343	100	343	0
02ae31010_34616666_f2_27	851	184	01cel11618orf1	503	124	99	125	60
02ae31010_34616666_f2_27	851	184	hp1p13947orf11	616	81	94	77	103
02ae31010_35270000_f3_33	852	231	01cp11504orf5	505	156	99	156	75
02ae31010_36132785_f2_29	853	438	01cel11618orf3	499	67	98	64	371
02ae31010_36132785_f2_29	853	438	01cel11618orf13	504	378	97	387	60
02ae31010_5085162_c1_47	934	416	07ce10203orf11	634	416	100	416	0
02cp10615_26573462_c1_45	935	168	hp3p10304orf2	727	126	91	129	42
02ge10116_15781452_c1_87	854	97	06ge10115orf17	563	97	100	97	0
02ge10116_16803513_f2_34	855	488	01cp11710orf16	652	64	100	62	424
02ge10116_16803513_f2_34	855	488	01cp11710orf5	654	146	99	145	342
02ge10116_36367936_c1_92	857	173	02ge20116orf25	521	173	100	173	0
03ae10804_12609533_c1_26	936	257	06ep10306orf10	610	257	100	257	0

- 49 -

hp6p10903_4398263_f3_6.aa	924	370	05gpl1901orf19	553	370	100	370	0
hp6p10903_4398263_f3_6.aa	924	370	05gpl1901orf19	553	370	100	370	0
03ae10804_21698400_c2_32	937	386	06ep10306orf11	611	386	100	386	0
03ae10804_23485968_c3_47	858	88	06ep10306orf5	561	88	100	88	0
04ep41903_26757937_f3_16	938	703	29ap10306orf3	711	325	97	319	378
04ep41903_26757937_f3_16	938	703	29ap11902orf1	712	280	92	280	423
04ep41903_4101593_f2_10.aa	940	1240	14ee21118orf1	706	365	96	358	875
05ce10910_25598277_f3_3	859	258	07ee11519orf1	567	170	96	166	88
05ep10815_26570332_c2_99	941	183	12ge10610orf2	702	97	98	97	86
05ep10815_4195292_c1_84	942	441	hp1e10554orf1	715	103	99	94	338
05ep10815_4719175_c1_83	943	663	06ee10207orf1	606	99	99	99	564
06ae11016_30579712_f2_21	860	239	03ce21717orf1	526	193	99	193	46
06ce20610_1367157_f1_8.aa	844	248	29ge10111orf3	714	124	90	121	124
06ce20610_29298537_c2_32	861	234	05gp20111orf4	555	234	100	234	0
06ce20610_3913967_c3_36	862	283	05gp20111orf6	556	283	100	283	0
06ce20610_4331338_f3_18	863	353	05gp20111orf12	554	243	99	245	110
06cp11118_212827_c1_17	864	73	06cp11118orf7	558	73	100	73	0
06cp11217_19720300_f3_11	865	308	hp1p13852orf6	614	646	96	278	-338
06cp11217_4881263_f2_9	866	93	hp1p13852orf7	615	93	100	93	0
06cp11217_4897077_f1_6	867	100	hp1p13852orf4	613	100	100	100	0
06cp30603_21492187_f2_41	868	106	04gpl1213orf14	539	136	98	106	-30
06cp30603_34024187_f1_20	869	483	04gpl1213orf11	538	122	100	122	361
06cp30603_34024187_f1_20	869	483	04gpl1213orf5	669	382	95	377	101
06cp30603_679218_f2_34	944	380	hp3e10302orf17	721	231	98	236	149
06ep10615_14649077_f3_52	871	300	04ce11617orf4	534	152	96	145	148
06ep10615_961562_f2_41	945	636	29ep20112orf2	713	135	98	87	501
06ep10615_9842_f3_46	872	159	05ep11717orf2	552	192	99	159	-33
06ep11202_133293_c1_19	946	167	06cp20302orf3	602	167	100	167	0
06ep11202_26353438_c1_22	873	286	hp2e11858orf6	623	155	100	154	131
06ep11202_26353438_c1_22	873	286	02ae21214orf1	522	109	92	110	177
06ep11202_792962_c2_26.aa	949	151	06cp20302orf6	604	66	92	66	85
06ep11202_4884677_c1_17	948	171	06cp20302orf5	603	171	99	171	0
06ep11917_24803153_c3_24	848	409	hp4e14535orf3	536	253	98	239	156
06ep11917_24803153_c3_24	848	409	hp4e14535orf7	730	60	97	60	349
06ep30223_16512_c3_160	951	231	hp3e10128orf1	720	231	100	231	0
06ep30223_23476067_c1_119	952	151	01ep30520orf8	511	94	93	95	57
06ep30223_23557202_c2_130	874	329	hp3e11024orf49	630	329	100	329	0
06ep30223_34409437_f3_94	875	148	hp3e11024orf5	631	148	100	148	0
06ep30223_4698838_f2_55	845	663	14gpl2015orf13	662	344	100	334	319
06ep30223_4698838_f2_55	845	663	14gpl2015orf16	663	117	99	112	546
06ep30223_4876077_c3_149	877	113	hp3e11024orf34	629	113	100	113	0
06ep30223_5109443_c1_109	878	342	05ee10411orf5	551	293	95	287	49
06ep30223_5271902_c1_106	879	207	hp2e10229orf4	619	80	95	38	127
06gp10409_3398427_f2_12	880	115	13ae10511orf3	583	115	92	115	0
06gp71906_15115637_f2_59	953	158	hp4p11393orf2	733	158	100	158	0
06gp71906_24261588_c2_174	881	96	01ae12021orf5	492	96	100	96	0
06gp71906_25478192_c1_131	954	236	hp3e11122orf3	723	236	100	236	0
06gp71906_25504187_f3_112	955	443	hp4p11393orf1	732	169	91	155	274
06gp71906_970325_c3_190	882	158	02ap21113orf2	512	113	92	98	45
07ae10923_24426508_f1_1	883	137	11ge10309orf28	685	137	100	137	0
07ae10923_24426508_f1_1	883	137	11ge10309orf28	685	137	100	137	0
09ce10413_414011_f1_3	885	169	02ce10216orf6	516	172	100	169	-3

- 50 -

09ce10413_5865665_f1_4	886	353	02ce10216orf7	517	74	100	74	279
09ce52017_29324062_c1_21	887	141	01ep30520orf24	506	141	100	141	0
09cp10224_1062966_c3_61	888	367	05ce10420orf1	548	194	100	194	173
09cp10224_1412715_c3_56	889	185	01ce10516orf20	495	185	100	185	0
09cp10224_429510_c2_46.aa	890	316	01ce10516orf21	496	108	98	108	208
09cp10224_4484718_c1_38	891	581	01ce10516orf15	493	301	99	297	280
09cp21607_7224187_c2_12	892	72	07gp31516orf9	569	72	100	72	0
09cp61003_14562637_c2_93	893	106	03gp20123orf3	527	80	98	83	26
09cp61003_19532625_c1_78	894	357	14ge10705orf14	599	357	99	357	0
09cp61003_24063587_c1_74	895	201	11ee10423orf1	577	116	100	116	85
09cp61003_24335762_c3_111	896	204	11ee10423orf4	578	105	100	104	99
09cp61003_492187_c2_80.aa	956	667	06gp10108orf2	621	266	98	266	401
09cp61003_5945252_f1_5	897	219	14ge10705orf3	695	219	100	219	0
11ae80818_11188791_c3_60	898	221	11ap11902orf3	575	185	99	185	36
11ae80818_19632781_c3_57	957	100	06cp11722orf11	595	100	100	100	0
11ae80818_7290627_c2_51	958	152	06cp11722orf5	598	98	100	93	54
11ae80818_783127_c3_63	899	182	03ap21820orf4	523	182	100	182	0
11ae80818_7952_c1_49	900	148	03ap21820orf8	524	148	100	148	0
11ap20714_34023312_f3_46	959	285	hp3e10057orf3	719	285	100	285	0
11ap20714_4960432_c3_97	901	151	14ee11217orf2	596	100	100	88	51
11ap20714_5271967_c1_60	902	260	14ee11217orf3	597	231	100	227	29
11ap20714_7227202_f3_43.aa	903	798	05ap21216orf4	547	529	100	529	269
11ap20714_7227202_f3_43.aa	903	798	05ap21216orf4	547	529	100	529	269
11ee11408_4977193_c1_41.aa	960	497	14ep11115orf3	708	91	100	91	406
11ge10308_5256_f2_1	961	71	11ge10308orf1	693	71	100	71	0
12ap10324_13178562_f3_6	962	269	12ap10324orf8	700	164	99	164	105
12ap10324_4805318_f2_3	846	326	12ap10324orf6	581	95	100	77	231
12ap10324_4805318_f2_3	846	326	12ap10324orf5	645	215	99	213	111
14ce31519_15635927_f3_15	905	293	02ce10809orf6	518	293	100	293	0
14ce61516_13073577_f2_12	963	784	14ep11905orf13	709	721	100	526	63
14ee41924_16282067_c1_72	964	64	02ep30607orf32	546	104	97	63	-40
14ee41924_23527267_c3_107	906	233	02ep30607orf27	520	138	98	97	95
14ee41924_23834800_f2_32	907	298	11ce10917orf10	639	298	100	298	0
14ee41924_2458267_c2_93	847	1304	02ep30607orf19	590	407	99	397	897
hp1p13939_25397327_f3_22	908	228	06ce10808orf2	557	151	100	145	77
hp2e10911_24855312_c1_69	909	161	11cp12006orf13	576	105	95	112	56
hp2e10911_3349_c1_63	910	113	01ce11618orf22	648	79	100	65	34
hp2e10911_4882027_c2_87	965	583	01cp11108orf6	507	101	99	99	482
hp3e11188_47327_f2_5	966	368	hp3p10807orf6	728	325	96	323	43
hp3e11188_5082842_f3_12	967	89	06ee11611orf1	608	89	100	89	0
hp4e13394_3368767_c1_80	968	804	09ap11406orf2	668	804	100	804	0
hp4e13394_35957200_f1_21	911	233	06ep11108orf17	562	165	100	147	68
hp4e13394_5964452_c2_97	912	79	02ap71220orf2	513	79	100	79	0
hp4e53394_22864682_c2_86.aa	913	427	hp1p13852orf6	614	646	100	422	-219
hp5e15044_4554652_f3_3	914	146	07cp10312orf5	677	138	100	138	8
hp5p15212_6928132_c3_34	969	270	11ce10908orf1	682	165	93	169	105
hp5p15575_29300311_c1_29	915	152	14ce10720orf3	591	163	95	152	-11
hp5p15575_33445317_f2_20.aa	916	294	hp3p11086orf1	636	217	92	202	77
hp5p15575_6140713_f2_18	917	288	14ce10720 rf12	589	288	100	288	0
hp5p15641_12195281_c1_24	918	92	hp5p15612orf2	643	92	100	92	0
hp5p15641_24304527_c3_35	919	139	29ge10307orf4	609	103	98	102	36
hp5p15641_25635452_c3_34	920	92	29ge10307orf3	607	92	100	92	0

- 51 -

hp5p15641_30273312_c2_28	970	301	05ce10613orf1	572	79	91	70	222
hp5p15641_30273312_c2_28	970	301	05ce10613orf2	573	61	100	54	240
hp5p15641_5211687_c2_29	971	297	05ap10914orf3	571	60	98	60	237
hp5p15870_14350428_f1_1	921	84	05ae20220orf56	543	101	99	79	-17
hp6p10590_23440913_c2_31	922	371	04ee70114orf10	535	243	98	235	128
hp6p10590_30521093_f2_14	972	138	07ce11206orf1	637	107	97	107	31
hp6p10606_19546933_c3_31	923	283	13ep12003orf21	587	222	92	228	61
hp6p10904_2214676_c1_14	926	179	hp4p13446orf3	640	179	99	179	0
hp6p10904_23704412_f2_5	927	384	hp4p13446orf13	638	384	99	384	0
hp6p10904_7089062_c1_16	973	364	hp4p13446orf5	736	364	100	364	0
hp6p12129_16603417_f3_14	974	358	hp3e10302orf26	722	268	98	266	90
hp6p12244_3948467_c1_52	975	476	hp4p12005orf2	735	108	96	84	368
hp6p22217_23470967_f1_4	976	272	13ee12016orf7	703	272	100	272	0
hp7e10192_4412568_f2_5	977	291	02ce10114orf3	525	291	100	291	0
hp7p10287_24611325_c2_24	978	256	01ce11513orf17	500	256	100	256	0
hp7p10290_25548812_f3_14	979	489	hp3p10156orf2	724	333	100	333	156
hp7p10290_25585941_f3_12	980	374	hp3p10156orf3	725	91	91	87	283
hp7p10290_35156558_f3_15	981	411	11cp12002orf3	689	109	96	107	302
hp7p10290_4351718_f1_6	982	324	hp3p10156orf7	726	302	95	299	22

Additional relationships between polypeptides of the invention are described in Table 5 below. All of the polypeptide lengths in Table 5 below are measured from start codon to stop codon in the nucleotide sequence of *H. pylori*.

5 The relationship between the polypeptides shown in Table 5 can be described as follows. First, all of the polypeptides of Table 5 are at least 90% identical with each other over most of their lengths, and most are over 95% identical with each other. Second, the start-to-stop lengths are different for some of the homologous pairs of polypeptides. In some cases, the shorter polypeptide contains the relevant portion of the protein exhibiting utility in this invention; in some cases, the longer polypeptide may exhibit improved utility. Third, some polypeptides in the second column are homologous to two shorter polypeptides in the fifth column.

10 In all cases, the homology relationships described in Table 5 are highly significant. For example, a typical *H. pylori* gene product will exhibit amino acid sequence identities of between 92% and 100% among different strains of *H. pylori* selected from human patients. The nucleotide sequences encoding the related polypeptides of this invention are also very similar to one another. For example, nucleotide probes derived from the coding sequence of a polypeptide of this invention can be used in PCR or hybridization experiments to identify clones carrying the nucleotide sequence encoding the homologous related polypeptide.

TABLE 5

ORF Name	aa Seq ID #	Length (aa)	ORF Name	aa Seq ID #	Length (aa)	Length Differ.	% Ident.	Overlap (aa)
01ce21104_33203250_c3_87	850	502	02ae11612_33203250_c1_51	752	502	0	99.8	502
02ge10116_15781452_c1_87	854	97	02gp20706_15781452_c2_51	756	80	17	100	80
02ge10116_16803513_f2_34	855	479	02gp20706_16803513_f1_1	802	486	-7	100	479
02ge10116_16803513_f2_34	855	479	07ep11916_5913592_f3_18	818	269	210	97.66	128
02ge10116_36367936_c1_92	857	171	02ge10116_36367936_c1_19	753	169	2	100	169
06cp30603_21492187_f2_41	868	89	hp6e20339_21492187_c1_40	795	89	0	97.75	89
06cp30603_23476568_c2_133.aa	843	437	06cp30603_23476568_c1_44	773	436	1	100	436
06cp30603_34024187_f1_20	869	479	hp6e20339_24317062_c3_57	833	357	122	100	357
06cp30603_34024187_f1_20	869	479	hp6e20339_34024187_c1_37	796	118	361	100	118
06ep10615_14649077_f3_52	871	300	05ee10816_14649077_f3_18	767	263	37	96.74	245
06gp1906_24261588_c2_174	881	89	04cp11202_24261588_f2_23	761	88	1	100	88
07ae10923_24426508_f1_1	883	133	06ap11119_24426508_f3_26	815	117	16	100	117
07ae10923_24426508_f1_1	883	133	29ep10720_24220926_f2_8	823	80	53	100	80
09cp61003_19532625_c1_78	894	352	09cp11003_19532625_c3_17	778	333	19	100	333
09cp61003_5945252_f1_5	897	211	09cp11003_5945252_f2_4	819	198	13	100	198
11ap20714_7227202_f3_43.aa	903	792	07ap20216_7227202_f3_10	776	533	259	100	527
14ee41924_2458267_c2_93	847	1288	07ap11015_23938312_f3_2	762	257	1031	100	257
hp2e10911_3349_c1_63	910	110	29ge30321_12913562_f1_1	825	70	40	100	65
hp3e11188_47327_f2_5	966	368	hp3p10807_189075_f2_4	838	312	56	100	312
hp5e15044_4554652_f3_3	914	114	hp6p10233_12273302_f1_1	834	143	-29	95.58	113
hp5p15870_14350428_f1_1	921	80	05ae30220_14350428_f1_9	811	748	-668	98.75	80
hp6p10903_4398263_f3_6.aa	924	366	01ae12001_24218781_f2_18	746	119	247	96.64	119
hp6p10903_4398263_f3_6.aa	924	366	14ap10221_13689381_c3_4	822	338	28	96.45	338
hp7p10287_24611325_c2_24	978	256	02ep20506_24611325_f2_6	835	242	14	100	242

- 53 -

Additional relationships between polypeptides of the invention are described in Table 6 below. All of the polypeptide lengths in Table 6 below are measured from stop codon to stop codon in the nucleotide sequence of *H. pylori*.

5 The relationship between the polypeptides shown in Table 6 can be described as follows. First, all of the polypeptides of Table 6 are at least 90% identical with each other over most of their lengths, and most are over 95% identical with each other. Second, the stop-to-stop lengths are different for the homologous pairs of polypeptides. In some cases, the shorter polypeptide contains the relevant portion of the protein exhibiting utility in this invention; in some cases, the longer polypeptide may exhibit  
10 improved utility.

In all cases, the homology relationships described in Table 6 are highly significant. For example, a typical *H. pylori* gene product will exhibit amino acid sequence identities of between 92% and 100% among different strains of *H. pylori* selected from human patients. The nucleotide sequences encoding the related  
15 polypeptides of this invention are also very similar to one another. For example, nucleotide probes derived from the coding sequence of a polypeptide of this invention can be used in PCR or hybridization experiments to identify clones carrying the nucleotide sequence encoding the homologous related polypeptide.

TABLE 6

ORF Name	aa Seq ID#	Leng.	ORF Name	aa Seq ID #	Leng.	Leng. Diff.	Overlap	% ID
01ae12001_2421871_f2_18	746	149	11ae11922_12386675_f2_1	1037	249	100	149	100.00
02gp20706_23866562_c2_53	758	120	02ge10116_23866562_c3_146	1038	641	521	102	100.00
07ap11015_23938312_f3_2	762	271	07ee11402_2458267_c3_108	1039	1304	1033	271	100.00
07ap11213_35401528_c1_21	764	803	07ee50709_10213593_f3_77	1040	1484	681	726	100.00
05ee10816_14649077_f3_18	767	263	06ep10615_14649077_f2_30	1041	300	37	245	96.74
06cp30603_23452_c3_80	772	177	09cp10713_23452_c3_195	1042	526	349	177	94.35
07ap20216_7227202_f3_10	776	533	11ap20714_7227202_f3_40	1043	798	265	533	100.00
09cp11003_19532625_c3_17	778	357	01ce61016_23609580_c3_139	1044	343	-14	342	100.00
12gp31106_3024126_f2_25	780	147	06gp71906_3024126_c1_128	1045	405	258	143	100.00
13ae10712_14100018_f2_12	781	201	06ge20501_14100018_c1_34	1046	553	352	201	100.00
hp6e20339_34024187_c1_37	796	122	09cp10713_34024187_f1_31	1047	483	361	122	100.00
01ge10203_35281542_c3_16	797	261	hp4p62853_5914693_c3_52	1048	272	11	261	99.62
04ge10816_22086531_f2_10	809	208	13ae10610_859692_c2_32	1049	401	193	208	98.56
04ge10816_33726080_c2_29	810	369	13ae10610_35912_f2_3	1050	459	90	361	100.00
06ap11119_24426508_f3_26	815	121	06ap11119_24426508_f3_27	1051	137	16	121	100.00
07ce11019_22051291_f1_1	817	600	14gp11423_26803801_f3_7	1052	712	112	568	99.82
14ap10221_13689381_c3_4	822	366	06ap10609_12586675_f2_19	1053	375	9	363	96.69
29ge30321_135253_f2_6	826	101	07ee50709_4818967_f2_43	1054	487	386	86	98.84
hp2p10272_22692325_f3_21	829	188	02ap11117_23495187_c3_81	1055	253	65	188	99.47
hp6e20339_24317062_c3_57	833	382	09cp10713_34024187_f1_31	1056	483	101	377	96.55
07ap11213_35156577_c1_24	836	401	07ee50709_35156577_f3_80	1057	567	166	400	100.00
07ap80601_976413_f3_9	837	274	05ae30220_976413_c3_204	1058	384	110	274	100.00
hp3p10807_189075_f2_4	838	312	hp3e11188_47327_f2_9	1059	368	56	312	100.00
12ap10324_4805318_f2_3	846	327	12ap10324_4805318_f2_6	1060	420	93	327	100.00
03ae10804_23485968_c3_47	858	88	hp1e80523_23485968_c2_49	1061	223	135	85	100.00
05ce10910_25598277_f3_3	859	258	hp7e10192_25598277_c2_15	1062	422	164	254	100.00
06cp11217_19720300_f3_11	865	308	hp4e53394_26209843_c3_98	1063	686	378	278	96.76
06ep10615_9842_f3_46	872	159	06ep10615_9842_f1_5	1064	345	186	159	99.37
06ep30223_34409437_f3_94	875	148	06ep30223_34409437_f2_64	1065	295	147	147	97.28

TABLE 6 (continued)

06gp10409_3398427_f2_12	880	116	06gp10409_3398427_f2_12	1066	292	176	116	100.00
09ce10413_5865665_f1_4	886	74	09ce10413_5865665_f1_4	1067	353	279	74	100.00
09cp10224_1062966_c3_61	888	367	09cp10224_1062966_c1_44	1068	485	118	372	94.09
09cp61003_14562637_c2_93	893	106	01ce61016_12931513_c2_106	1069	376	270	94	93.62
09cp61003_19532625_c1_78	894	357	01ce61016_23609580_c3_139	1070	343	-14	342	100.00
11ae80818_11188791_c3_60	898	221	14cp11908_25593768_c3_97	1071	538	317	204	99.51
11ae80818_783127_c3_63	899	182	14cp11908_783127_c1_72	1072	337	155	167	98.20
14ee41924_23527267_c3_107	906	233	07ee11402_10759567_c2_86	1073	292	59	232	100.00
hp2e10911_3349_c1_63	910	113	07ee50709_4818967_f2_43	1074	487	374	109	95.41
hp4e13394_35957200_f1_21	911	233	hp4e13394_5088562_f3_54	1075	225	-8	215	100.00
hp4e13394_5964452_c2_97	912	79	hp4e13394_15828963_c2_90	1076	135	56	77	100.00
hp4e53394_22864682_c2_86.aa	913	427	hp4e53394_19720300_c3_98	1077	647	220	422	100.00
hp5e15044_4554652_f3_3	914	146	07ce10312_4554652_f3_2	1078	174	28	146	100.00
hp5p15870_14350428_f1_1	921	84	05ae30220_14350428_f3_91	1079	752	668	84	98.81
hp6p10606_19546933_c3_31	923	283	hp8e10080_19546933_c2_88	1080	428	145	280	99.29
01cp20708_36134808_f2_11	928	230	01ce10320_30273587_f3_38	1081	275	45	230	100.00
02ae31010_2117087_f3_34	931	137	07ee50709_26438968_f2_36	1082	265	128	128	98.44
05ep10815_4719175_c1_83	943	663	05ep10815_4719175_c1_115	1083	925	262	653	100.00
06ep30223_23476067_c1_119	952	151	06ep30223_23476067_c1_115	1084	210	59	150	98.67
11ae80818_7290627_c2_51	958	152	hp7e10590_26172564_c1_68	1085	356	204	141	88.65
11ee11408_4977193_c1_41.aa	960	530	05ae30220_4977193_c3_198	1086	531	1	528	99.05
11ge10308_5256_f2_1	961	71	hp7e10557_21698387_f1_1	1087	239	168	69	98.55
14ee41924_16282067_c1_72	964	64	07ee11402_19565702_c2_88	1088	465	401	63	100.00
hp2e10911_4882027_c2_87	965	583	07ee50709_960952_f2_47	1089	1213	630	585	97.78
hp6p12244_3948467_c1_52	975	476	hp6p12244_3948467_c3_88	1090	508	32	445	100.00
14ce61516_13073577_f2_12	963	785	hp7e10590_13073577_c3_107	1296	897	112	785	100.00
09cp61003_492187_c2_80	956	667	01ce61016_492187_c3_120	1297	668	1	667	100.00
06ep10615_961562_f2_41	945	636	06ep10615_961562_f1_15	1298	637	1	636	100.00

#### IV. Identification of Nucleic Acids Encoding Vaccine Components and Targets for Agents Effective Against *H. pylori*

The disclosed *H. pylori* genome sequence includes segments that direct the synthesis of ribonucleic acids and polypeptides, as well as origins of replication, promoters, other types of regulatory sequences, and intergenic nucleic acids. The invention encompasses the identification of nucleic acids encoding immunogenic components of vaccines and targets for agents effective against *H. pylori*. An important aspect of this identification is to determine the function of the disclosed sequences, which can be achieved using a variety of approaches. Non-limiting examples of these methods are described briefly below.

Homology to known sequences: Computer-assisted comparison of the disclosed *H. pylori* sequences with previously reported sequences present in publicly available databases is a useful tool for identifying functional *H. pylori* nucleic acid and polypeptide sequences. It will be understood that protein-coding sequences, for example, may be compared as a whole, and that a high degree of sequence homology between two proteins (such as, for example, >80-90%) at the amino acid level is strongly suggestive that the two proteins also possess some degree of functional homology, such as, for example, among enzymes involved in metabolism, DNA synthesis, or cell wall synthesis, and proteins involved in transport, cell division, etc. In addition, many structural features of particular protein classes have been identified and correlate with specific consensus sequences, such as, for example, binding domains for nucleotides, DNA, metal ions, and other small molecules; sites for covalent modifications such as phosphorylation, acylation, and the like; sites of protein:protein interactions, etc. These consensus sequences may be quite short and thus may represent only a fraction of the entire protein-coding sequence. Identification of such a feature in an *H. pylori* sequence is therefore useful in determining the function of the encoded protein and identifying potentially useful targets of antibacterial drugs.

Of particular relevance to the present invention are structural features that are common to secretory, transmembrane, and surface proteins, including secretion signal peptides and hydrophobic transmembrane domains. *H. pylori* proteins identified as containing putative signal sequences and/or transmembrane domains are useful as immunogenic components of vaccines.

Identification of essential genes: Nucleic acids that encode proteins essential for growth or viability of *H. pylori* are preferred drug targets. *H. pylori* genes can be tested for their biological relevance to the organism by examining the effect of deleting and/or disrupting the genes, i.e., by so-called gene "knockout", using techniques known to those skilled in the relevant art. In this manner, essential genes may be identified.

- 57 -

Strain-specific sequences: Because of the evolutionary relationship between different *H. pylori* strains, it is believed that the presently disclosed *H. pylori* sequences are useful for identifying, and/or discriminating between, previously known and new *H. pylori* strains. It is believed that other *H. pylori* strains will exhibit at least 70% sequence homology with the presently disclosed sequence, although whether or not this is correct is not essential to the invention. Systematic and routine analyses of DNA sequences derived from samples containing *H. pylori* strains, and comparison with the present sequence allows for the identification of sequences that can be used to discriminate between strains, as well as those that are common to all *H. pylori* strains.

5 In one embodiment, the invention provides nucleic acids, including probes, and peptide and polypeptide sequences that discriminate between different strains of *H. pylori*. Strain-specific components can also be identified functionally by their ability to elicit or react with antibodies that selectively recognize one or more *H. pylori* strains.

10

In another embodiment, the invention provides nucleic acids, including probes, and peptide and polypeptide sequences that are common to all *H. pylori* strains but are not found in other bacterial species.

15

Specific Example: Determination Of Candidate Protein Antigens For Antibody And Vaccine Development

20 The selection of candidate protein antigens for vaccine development can be derived from the nucleic acids encoding *H. pylori* polypeptides. First, the ORF's can be analyzed for homology to other known exported or membrane proteins and analyzed using the discriminant analysis described by Klein, et al. (Klein, P., Kaneshia, M., and DeLisi, C. (1985) *Biochimica et Biophysica Acta* 815, 468-476) for predicting exported and membrane proteins.

25

Homology searches can be performed using the BLAST algorithm contained in the Wisconsin Sequence Analysis Package (Genetics Computer Group, University Research Park, 575 Science Drive, Madison, WI 53711) to compare each predicted ORF amino acid sequence with all sequences found in the current GenBank, SWISS-PROT and PIR databases. BLAST searches for local alignments between the ORF and the databank sequences and reports a probability score which indicates the probability of finding this sequence by chance in the database. ORF's with significant homology (e.g. probabilities better than  $1 \times 10^{-6}$ ) to membrane or exported proteins represent likely protein antigens for vaccine development. Possible functions can be provided to *H. pylori* genes based on sequence homology to genes cloned in other organisms.

30

35

Discriminant analysis (Klein, et al. supra) can be used to examine the ORF amino acid sequences. This algorithm uses the intrinsic information contained in the

- 58 -

ORF amino acid sequence and compares it to information derived from the properties of known membrane and exported proteins. This comparison predicts which proteins will be exported, membrane associated or cytoplasmic. ORF amino acid sequences identified as exported or membrane associated by this algorithm are likely protein  
5 antigens for vaccine development.

Surface exposed outer membrane proteins are likely to represent the best antigens to provide a protective immune response against *H. pylori*. Among the algorithms that can be used to aid in prediction of these outer membrane proteins include the presence of an amphipathic beta-sheet region at their C-terminus. This region which  
10 has been detected in a large number of outer membrane proteins in Gram negative bacteria is often characterized by hydrophobic residues (Phe or Tyr) approximately at positions 1, 3, 5, 7 and 9 from the C-terminus (e.g., see Figure 8, block F). Importantly, these sequences have not been detected at the C-termini of periplasmic proteins, thus allowing preliminary distinction between these classes of proteins based on primary  
15 sequence data. This phenomenon has been reported previously by Struyve et al. (*J. Mol. Biol.* 218:141-148, 1991).

Also illustrated in Figure 8 are additional amino acid sequence motifs found in many outer membrane proteins of *H. pylori*. The amino acid sequence alignment in Figure 8 depicts portions of the sequence of 12 *H. pylori* proteins (depicted in the single  
20 letter amino acid code) labeled with their amino acid Sequence ID Numbers and shown N-terminal to C-terminal, left to right. Six distinct blocks (labeled A through F) of similar amino acid residues are found including the distinctive hydrophobic residues (Phe or Tyr; F or Y according to the single letter code for amino acid residues) frequently found at positions near the C-terminus of outer membrane proteins. The  
25 presence of several shared motifs clearly establishes the similarity between members of this group of proteins.

In addition, outer membrane proteins isolated from *H. pylori* frequently share a motif near the mature N-terminus (i.e., after processing to remove the secretion signal) as illustrated in the blocked amino acid residues in Figure 9. Figure 9 depicts the N-  
30 terminal portion of nine *H. pylori* proteins (designated by their amino acid Sequence ID Numbers and shown N-terminal to C-terminal, left to right).

One skilled in the art would know that these shared sequence motifs are highly significant and establish a similarity among this group of proteins.

Infrequently it is not possible to distinguish between multiple possible  
35 nucleotides at a given position in the nucleic acid sequence. In those cases the ambiguities are denoted by an extended alphabet as follows:

- 59 -

These are the official IUPAC-IUB single-letter base codes

Code	Base Description	
G	Guanine	
A	Adenine	
T	Thymine	
C	Cytosine	
R	Purine	(A or G)
Y	Pyrimidine	(C or T or U)
M	Amino	(A or C)
K	Ketone	(G or T)
S	Strong interaction	(C or G)
W	Weak interaction	(A or T)
H	Not-G	(A or C or T)
B	Not-A	(C or G or T)
V	Not-T (not-U)	(A or C or G)
D	Not-C	(A or G or T)
N	Any	(A or C or G or T)

- 5 The amino acid translations of this invention account for the ambiguity in the nucleic acid sequence by translating the ambiguous codon as the letter "X". In all cases, the permissible amino acid residues at a position are clear from an examination of the nucleic acid sequence based on the standard genetic code.

#### V. Production of Fragments and Analogs of *H. pylori* Nucleic Acids and Polypeptides

- 10 Based on the discovery of the *H. pylori* gene products provided in the Sequence Listing, one skilled in the art can alter the disclosed structure (of *H. pylori* genes), e.g., by producing fragments or analogs, and test the newly produced structures for activity. Examples of techniques known to those skilled in the relevant art which allow the production and testing of fragments and analogs are discussed below. These, or  
15 analogous methods can be used to make and screen libraries of polypeptides, e.g., libraries of random peptides or libraries of fragments or analogs of cellular proteins for the ability to bind *H. pylori* polypeptides. Such screens are useful for discovery of inhibitors of *H. pylori*.

##### Generation of Fragments

- 20 Fragments of a protein can be produced in several ways, e.g., recombinantly, by proteolytic digestion, or by chemical synthesis. Internal or terminal fragments of a polypeptide can be generated by removing one or more nucleotides from one end (for a terminal fragment) or both ends (for an internal fragment) of a nucleic acid which encodes the polypeptide. Expression of the mutagenized DNA produces polypeptide  
25 fragments. Digestion with "end-nibbling" endonucleases can thus generate DNA's which encode an array of fragments. DNA's which encode fragments of a protein can

- 60 -

also be generated by random shearing, restriction digestion or a combination of the above-discussed methods.

Fragments can also be chemically synthesized using techniques known in the art such as conventional Merrifield solid phase f-Moc or t-Boc chemistry. For example, peptides of the present invention may be arbitrarily divided into fragments of desired length with no overlap of the fragments, or divided into overlapping fragments of a desired length.

#### Alteration of Nucleic Acids and Polypeptides: Random Methods

Amino acid sequence variants of a protein can be prepared by random mutagenesis of DNA which encodes a protein or a particular domain or region of a protein. Useful methods include PCR mutagenesis and saturation mutagenesis. A library of random amino acid sequence variants can also be generated by the synthesis of a set of degenerate oligonucleotide sequences. (Methods for screening proteins in a library of variants are elsewhere herein).

##### (A) PCR Mutagenesis

In PCR mutagenesis, reduced Taq polymerase fidelity is used to introduce random mutations into a cloned fragment of DNA (Leung et al., 1989, *Technique* 1:11-15). The DNA region to be mutagenized is amplified using the polymerase chain reaction (PCR) under conditions that reduce the fidelity of DNA synthesis by Taq DNA polymerase, e.g., by using a dGTP/dATP ratio of five and adding  $Mn^{2+}$  to the PCR reaction. The pool of amplified DNA fragments are inserted into appropriate cloning vectors to provide random mutant libraries.

##### (B) Saturation Mutagenesis

Saturation mutagenesis allows for the rapid introduction of a large number of single base substitutions into cloned DNA fragments (Mayers et al., 1985, *Science* 229:242). This technique includes generation of mutations, e.g., by chemical treatment or irradiation of single-stranded DNA *in vitro*, and synthesis of a complimentary DNA strand. The mutation frequency can be modulated by modulating the severity of the treatment, and essentially all possible base substitutions can be obtained. Because this procedure does not involve a genetic selection for mutant fragments both neutral substitutions, as well as those that alter function, are obtained. The distribution of point mutations is not biased toward conserved sequence elements.

##### (C) Degenerate Oligonucleotides

A library of homologs can also be generated from a set of degenerate oligonucleotide sequences. Chemical synthesis of a degenerate sequences can be carried out in an automatic DNA synthesizer, and the synthetic genes then ligated into an

- 61 -

appropriate expression vector. The synthesis of degenerate oligonucleotides is known in the art (see for example, Narang, SA (1983) *Tetrahedron* 39:3; Itakura et al. (1981) *Recombinant DNA, Proc 3rd Cleveland Sympos. Macromolecules*, ed. AG Walton, Amsterdam: Elsevier pp273-289; Itakura et al. (1984) *Annu. Rev. Biochem.* 53:323; Itakura et al. (1984) *Science* 198:1056; Ike et al. (1983) *Nucleic Acid Res.* 11:477. Such techniques have been employed in the directed evolution of other proteins (see, for example, Scott et al. (1990) *Science* 249:386-390; Roberts et al. (1992) *PNAS* 89:2429-2433; Devlin et al. (1990) *Science* 249: 404-406; Cwirla et al. (1990) *PNAS* 87: 6378-6382; as well as U.S. Patents Nos. 5,223,409, 5,198,346, and 5,096,815).

#### 10      Alteration of Nucleic Acids and Polypeptides: Methods for Directed Mutagenesis

Non-random or directed, mutagenesis techniques can be used to provide specific sequences or mutations in specific regions. These techniques can be used to create variants which include, e.g., deletions, insertions, or substitutions, of residues of the known amino acid sequence of a protein. The sites for mutation can be modified individually or in series, e.g., by (1) substituting first with conserved amino acids and then with more radical choices depending upon results achieved, (2) deleting the target residue, or (3) inserting residues of the same or a different class adjacent to the located site, or combinations of options 1-3.

##### (A) Alanine Scanning Mutagenesis

20      Alanine scanning mutagenesis is a useful method for identification of certain residues or regions of the desired protein that are preferred locations or domains for mutagenesis, Cunningham and Wells (*Science* 244:1081-1085, 1989). In alanine scanning, a residue or group of target residues are identified (e.g., charged residues such as Arg, Asp, His, Lys, and Glu) and replaced by a neutral or negatively charged amino acid (most preferably alanine or polyalanine). Replacement of an amino acid can affect the interaction of the amino acids with the surrounding aqueous environment in or outside the cell. Those domains demonstrating functional sensitivity to the substitutions are then refined by introducing further or other variants at or for the sites of substitution. Thus, while the site for introducing an amino acid sequence variation is predetermined, the nature of the mutation per se need not be predetermined. For example, to optimize the performance of a mutation at a given site, alanine scanning or random mutagenesis may be conducted at the target codon or region and the expressed desired protein subunit variants are screened for the optimal combination of desired activity.

##### (B) Oligonucleotide-Mediated Mutagenesis

35      Oligonucleotide-mediated mutagenesis is a useful method for preparing substitution, deletion, and insertion variants of DNA, see, e.g., Adelman et al., (*DNA*

- 62 -

2:183, 1983). Briefly, the desired DNA is altered by hybridizing an oligonucleotide encoding a mutation to a DNA template, where the template is the single-stranded form of a plasmid or bacteriophage containing the unaltered or native DNA sequence of the desired protein. After hybridization, a DNA polymerase is used to synthesize an entire  
5 second complementary strand of the template that will thus incorporate the oligonucleotide primer, and will code for the selected alteration in the desired protein DNA. Generally, oligonucleotides of at least 25 nucleotides in length are used. An optimal oligonucleotide will have 12 to 15 nucleotides that are completely complementary to the template on either side of the nucleotide(s) coding for the  
10 mutation. This ensures that the oligonucleotide will hybridize properly to the single-stranded DNA template molecule. The oligonucleotides are readily synthesized using techniques known in the art such as that described by Crea et al. (*Proc. Natl. Acad. Sci. USA*, 75: 5765[1978]).

#### (C) Cassette Mutagenesis

15 Another method for preparing variants, cassette mutagenesis, is based on the technique described by Wells et al. (*Gene*, 34:315[1985]). The starting material is a plasmid (or other vector) which includes the protein subunit DNA to be mutated. The codon(s) in the protein subunit DNA to be mutated are identified. There must be a unique restriction endonuclease site on each side of the identified mutation site(s). If no  
20 such restriction sites exist, they may be generated using the above-described oligonucleotide-mediated mutagenesis method to introduce them at appropriate locations in the desired protein subunit DNA. After the restriction sites have been introduced into the plasmid, the plasmid is cut at these sites to linearize it. A double-stranded oligonucleotide encoding the sequence of the DNA between the restriction sites but  
25 containing the desired mutation(s) is synthesized using standard procedures. The two strands are synthesized separately and then hybridized together using standard techniques. This double-stranded oligonucleotide is referred to as the cassette. This cassette is designed to have 3' and 5' ends that are comparable with the ends of the linearized plasmid, such that it can be directly ligated to the plasmid. This plasmid now  
30 contains the mutated desired protein subunit DNA sequence.

#### (D) Combinatorial Mutagenesis

Combinatorial mutagenesis can also be used to generate mutants (Ladner et al., WO 88/06630). In this method, the amino acid sequences for a group of homologs or other related proteins are aligned, preferably to promote the highest homology  
35 possible. All of the amino acids which appear at a given position of the aligned sequences can be selected to create a degenerate set of combinatorial sequences. The

- 63 -

variegated library of variants is generated by combinatorial mutagenesis at the nucleic acid level, and is encoded by a variegated gene library. For example, a mixture of synthetic oligonucleotides can be enzymatically ligated into gene sequences such that the degenerate set of potential sequences are expressible as individual peptides, or  
5 alternatively, as a set of larger fusion proteins containing the set of degenerate sequences.

#### Other Modifications of *H. pylori* Nucleic Acids and Polypeptides

It is possible to modify the structure of an *H. pylori* polypeptide for such purposes as increasing solubility, enhancing stability (e.g., shelf life *ex vivo* and  
10 resistance to proteolytic degradation *in vivo*). A modified *H. pylori* protein or peptide can be produced in which the amino acid sequence has been altered, such as by amino acid substitution, deletion, or addition as described herein.

An *H. pylori* peptide can also be modified by substitution of cysteine residues preferably with alanine, serine, threonine, leucine or glutamic acid residues to minimize  
15 dimerization via disulfide linkages. In addition, amino acid side chains of fragments of the protein of the invention can be chemically modified. Another modification is cyclization of the peptide.

In order to enhance stability and/or reactivity, an *H. pylori* polypeptide can be modified to incorporate one or more polymorphisms in the amino acid sequence of the  
20 protein resulting from any natural allelic variation. Additionally, D-amino acids, non-natural amino acids, or non-amino acid analogs can be substituted or added to produce a modified protein within the scope of this invention. Furthermore, an *H. pylori* polypeptide can be modified using polyethylene glycol (PEG) according to the method of A. Sehon and co-workers (Wie et al., *supra*) to produce a protein conjugated with  
25 PEG. In addition, PEG can be added during chemical synthesis of the protein. Other modifications of *H. pylori* proteins include reduction/alkylation (Tarr, *Methods of Protein Microcharacterization*, J. E. Silver ed., Humana Press, Clifton NJ 155-194 (1986)); acylation (Tarr, *supra*); chemical coupling to an appropriate carrier (Mishell and Shiigi, eds, *Selected Methods in Cellular Immunology*, WH Freeman, San Francisco, CA  
30 (1980), U.S. Patent 4,939,239; or mild formalin treatment (Marsh, (1971) *Int. Arch. of Allergy and Appl. Immunol.*, 41: 199 - 215).

To facilitate purification and potentially increase solubility of an *H. pylori* protein or peptide, it is possible to add an amino acid fusion moiety to the peptide backbone. For example, hexa-histidine can be added to the protein for purification by  
35 immobilized metal ion affinity chromatography (Hochuli, E. et al., (1988) *Bio/Technology*, 6: 1321 - 1325). In addition, to facilitate isolation of peptides free of

- 64 -

irrelevant sequences, specific endoprotease cleavage sites can be introduced between the sequences of the fusion moiety and the peptide.

To potentially aid proper antigen processing of epitopes within an *H. pylori* polypeptide, canonical protease sensitive sites can be engineered between regions, each comprising at least one epitope via recombinant or synthetic methods. For example, charged amino acid pairs, such as KK or RR, can be introduced between regions within a protein or fragment during recombinant construction thereof. The resulting peptide can be rendered sensitive to cleavage by cathepsin and/or other trypsin-like enzymes which would generate portions of the protein containing one or more epitopes. In addition, such charged amino acid residues can result in an increase in the solubility of the peptide.

#### Primary Methods for Screening Polypeptides and Analogs

Various techniques are known in the art for screening generated mutant gene products. Techniques for screening large gene libraries often include cloning the library into replicable expression vectors, transforming appropriate cells with the resulting library of vectors, and expressing the genes under conditions in which detection of a desired activity, e.g., in this case, binding to *H. pylori* polypeptide or an interacting protein, facilitates relatively easy isolation of the vector encoding the gene whose product was detected. Each of the techniques described below is amenable to high through-put analysis for screening large numbers of sequences created, e.g., by random mutagenesis techniques.

#### (A) Two Hybrid Systems

Two hybrid assays such as the system described above (as with the other screening methods described herein), can be used to identify polypeptides, e.g., fragments or analogs of a naturally-occurring *H. pylori* polypeptide, e.g., of cellular proteins, or of randomly generated polypeptides which bind to an *H. pylori* protein. (The *H. pylori* domain is used as the bait protein and the library of variants are expressed as fish fusion proteins.) In an analogous fashion, a two hybrid assay (as with the other screening methods described herein), can be used to find polypeptides which bind a *H. pylori* polypeptide.

#### (B) Display Libraries

In one approach to screening assays, the candidate peptides are displayed on the surface of a cell or viral particle, and the ability of particular cells or viral particles to bind an appropriate receptor protein via the displayed product is detected in a "panning assay". For example, the gene library can be cloned into the gene for a surface membrane protein of a bacterial cell, and the resulting fusion protein detected by

- 65 -

panning (Ladner et al., WO 88/06630; Fuchs et al. (1991) *Bio/Technology* 9:1370-1371; and Goward et al. (1992) *TIBS* 18:136-140). In a similar fashion, a detectably labeled ligand can be used to score for potentially functional peptide homologs. Fluorescently labeled ligands, e.g., receptors, can be used to detect homolog which retain ligand-binding activity. The use of fluorescently labeled ligands, allows cells to be visually inspected and separated under a fluorescence microscope, or, where the morphology of the cell permits, to be separated by a fluorescence-activated cell sorter.

A gene library can be expressed as a fusion protein on the surface of a viral particle. For instance, in the filamentous phage system, foreign peptide sequences can be expressed on the surface of infectious phage, thereby conferring two significant benefits. First, since these phage can be applied to affinity matrices at concentrations well over  $10^{13}$  phage per milliliter, a large number of phage can be screened at one time. Second, since each infectious phage displays a gene product on its surface, if a particular phage is recovered from an affinity matrix in low yield, the phage can be amplified by another round of infection. The group of almost identical *E. coli* filamentous phages M13, fd., and f1 are most often used in phage display libraries. Either of the phage gIII or gVIII coat proteins can be used to generate fusion proteins without disrupting the ultimate packaging of the viral particle. Foreign epitopes can be expressed at the NH<sub>2</sub>-terminal end of pIII and phage bearing such epitopes recovered from a large excess of phage lacking this epitope (Ladner et al. PCT publication WO 90/02909; Garrard et al., PCT publication WO 92/09690; Marks et al. (1992) *J. Biol. Chem.* 267:16007-16010; Griffiths et al. (1993) *EMBO J* 12:725-734; Clackson et al. (1991) *Nature* 352:624-628; and Barbas et al. (1992) *PNAS* 89:4457-4461).

A common approach uses the maltose receptor of *E. coli* (the outer membrane protein, LamB) as a peptide fusion partner (Charbit et al. (1986) *EMBO* 5, 3029-3037). Oligonucleotides have been inserted into plasmids encoding the LamB gene to produce peptides fused into one of the extracellular loops of the protein. These peptides are available for binding to ligands, e.g., to antibodies, and can elicit an immune response when the cells are administered to animals. Other cell surface proteins, e.g., OmpA (Schorr et al. (1991) *Vaccines* 91, pp. 387-392), PhoE (Agterberg, et al. (1990) *Gene* 88, 37-45), and PAL (Fuchs et al. (1991) *Bio/Tech* 9, 1369-1372), as well as large bacterial surface structures have served as vehicles for peptide display. Peptides can be fused to pilin, a protein which polymerizes to form the pilus-a conduit for interbacterial exchange of genetic information (Thiry et al. (1989) *Appl. Environ. Microbiol.* 55, 984-993). Because of its role in interacting with other cells, the pilus provides a useful support for the presentation of peptides to the extracellular environment. Another large surface

- 66 -

structure used for peptide display is the bacterial motive organ, the flagellum. Fusion of peptides to the subunit protein flagellin offers a dense array of many peptides copies on the host cells (Kuwajima et al. (1988) *Bio/Tech.* 6, 1080-1083). Surface proteins of other bacterial species have also served as peptide fusion partners. Examples include the

5 *Staphylococcus* protein A and the outer membrane protease IgA of *Neisseria* (Hansson et al. (1992) *J. Bacteriol.* 174, 4239-4245 and Klauser et al. (1990) *EMBO J.* 9, 1991-1999).

In the filamentous phage systems and the LamB system described above, the physical link between the peptide and its encoding DNA occurs by the containment of the DNA within a particle (cell or phage) that carries the peptide on its surface.

10 Capturing the peptide captures the particle and the DNA within. An alternative scheme uses the DNA-binding protein LacI to form a link between peptide and DNA (Cull et al. (1992) *PNAS USA* 89:1865-1869). This system uses a plasmid containing the LacI gene with an oligonucleotide cloning site at its 3'-end. Under the controlled induction by

15 arabinose, a LacI-peptide fusion protein is produced. This fusion retains the natural ability of LacI to bind to a short DNA sequence known as LacO operator (LacO). By installing two copies of LacO on the expression plasmid, the LacI-peptide fusion binds tightly to the plasmid that encoded it. Because the plasmids in each cell contain only a single oligonucleotide sequence and each cell expresses only a single peptide sequence,

20 the peptides become specifically and stably associated with the DNA sequence that directed its synthesis. The cells of the library are gently lysed and the peptide-DNA complexes are exposed to a matrix of immobilized receptor to recover the complexes containing active peptides. The associated plasmid DNA is then reintroduced into cells for amplification and DNA sequencing to determine the identity of the peptide ligands.

25 As a demonstration of the practical utility of the method, a large random library of dodecapeptides was made and selected on a monoclonal antibody raised against the opioid peptide dynorphin B. A cohort of peptides was recovered, all related by a consensus sequence corresponding to a six-residue portion of dynorphin B. (Cull et al. (1992) *Proc. Natl. Acad. Sci. U.S.A.* 89-1869)

30 This scheme, sometimes referred to as peptides-on-plasmids, differs in two important ways from the phage display methods. First, the peptides are attached to the C-terminus of the fusion protein, resulting in the display of the library members as peptides having free carboxy termini. Both of the filamentous phage coat proteins, pIII and pVIII, are anchored to the phage through their C-termini, and the guest peptides are

35 placed into the outward-extending N-terminal domains. In some designs, the phage-displayed peptides are presented right at the amino terminus of the fusion protein.

- 67 -

(Cwirla, et al. (1990) *Proc. Natl. Acad. Sci. U.S.A.* 87, 6378-6382) A second difference is the set of biological biases affecting the population of peptides actually present in the libraries. The LacI fusion molecules are confined to the cytoplasm of the host cells. The phage coat fusions are exposed briefly to the cytoplasm during translation but are rapidly secreted through the inner membrane into the periplasmic compartment, remaining anchored in the membrane by their C-terminal hydrophobic domains, with the N-termini, containing the peptides, protruding into the periplasm while awaiting assembly into phage particles. The peptides in the LacI and phage libraries may differ significantly as a result of their exposure to different proteolytic activities. The phage coat proteins require transport across the inner membrane and signal peptidase processing as a prelude to incorporation into phage. Certain peptides exert a deleterious effect on these processes and are underrepresented in the libraries (Gallop et al. (1994) *J. Med. Chem.* 37(9):1233-1251). These particular biases are not a factor in the LacI display system.

The number of small peptides available in recombinant random libraries is enormous. Libraries of  $10^7$ - $10^9$  independent clones are routinely prepared. Libraries as large as  $10^{11}$  recombinants have been created, but this size approaches the practical limit for clone libraries. This limitation in library size occurs at the step of transforming the DNA containing randomized segments into the host bacterial cells. To circumvent this limitation, an *in vitro* system based on the display of nascent peptides in polysome complexes has recently been developed. This display library method has the potential of producing libraries 3-6 orders of magnitude larger than the currently available phage/phagemid or plasmid libraries. Furthermore, the construction of the libraries, expression of the peptides, and screening, is done in an entirely cell-free format.

In one application of this method (Gallop et al. (1994) *J. Med. Chem.* 37(9):1233-1251), a molecular DNA library encoding  $10^{12}$  decapeptides was constructed and the library expressed in an *E. coli* S30 *in vitro* coupled transcription/translation system. Conditions were chosen to stall the ribosomes on the mRNA, causing the accumulation of a substantial proportion of the RNA in polysomes and yielding complexes containing nascent peptides still linked to their encoding RNA. The polysomes are sufficiently robust to be affinity purified on immobilized receptors in much the same way as the more conventional recombinant peptide display libraries are screened. RNA from the bound complexes is recovered, converted to cDNA, and amplified by PCR to produce a template for the next round of synthesis and screening. The polysome display method can be coupled to the phage display system. Following several rounds of screening, cDNA from the enriched pool of polysomes was cloned into

- 68 -

a phagemid vector. This vector serves as both a peptide expression vector, displaying peptides fused to the coat proteins, and as a DNA sequencing vector for peptide identification. By expressing the polysome-derived peptides on phage, one can either continue the affinity selection procedure in this format or assay the peptides on individual clones for binding activity in a phage ELISA, or for binding specificity in a completion phage ELISA (Barret, et al. (1992) *Anal. Biochem* 204,357-364). To identify the sequences of the active peptides one sequences the DNA produced by the phagemid host.

#### Secondary Screening of Polypeptides and Analogs

The high through-put assays described above can be followed by secondary screens in order to identify further biological activities which will, e.g., allow one skilled in the art to differentiate agonists from antagonists. The type of a secondary screen used will depend on the desired activity that needs to be tested. For example, an assay can be developed in which the ability to inhibit an interaction between a protein of interest and its respective ligand can be used to identify antagonists from a group of peptide fragments isolated though one of the primary screens described above.

Therefore, methods for generating fragments and analogs and testing them for activity are known in the art. Once the core sequence of interest is identified, it is routine to perform for one skilled in the art to obtain analogs and fragments.

#### Peptide Mimetics of *H. pylori* Polypeptides

The invention also provides for reduction of the protein binding domains of the subject *H. pylori* polypeptides to generate mimetics, e.g. peptide or non-peptide agents. The peptide mimetics are able to disrupt binding of a polypeptide to its counter ligand, e.g., in the case of an *H. pylori* polypeptide binding to a naturally occurring ligand. The critical residues of a subject *H. pylori* polypeptide which are involved in molecular recognition of a polypeptide can be determined and used to generate *H. pylori*-derived peptidomimetics which competitively or noncompetitively inhibit binding of the *H. pylori* polypeptide with an interacting polypeptide (see, for example, European patent applications EP-412,762A and EP-B31,080A).

For example, scanning mutagenesis can be used to map the amino acid residues of a particular *H. pylori* polypeptide involved in binding an interacting polypeptide, peptidomimetic compounds (e.g. diazepam or isoquinoline derivatives) can be generated which mimic those residues in binding to an interacting polypeptide, and which therefore can inhibit binding of an *H. pylori* polypeptide to an interacting polypeptide and thereby interfere with the function of *H. pylori* polypeptide. For instance, non-hydrolyzable peptide analogs of such residues can be generated using benzodiazepine

- 69 -

(e.g., see Freidinger et al. in *Peptides: Chemistry and Biology*, G.R. Marshall ed., ESCOM Publisher: Leiden, Netherlands, 1988), azepine (e.g., see Huffman et al. in *Peptides: Chemistry and Biology*, G.R. Marshall ed., ESCOM Publisher: Leiden, Netherlands, 1988), substituted gamma lactam rings (Garvey et al. in *Peptides: Chemistry and Biology*, G.R. Marshall ed., ESCOM Publisher: Leiden, Netherlands, 1988), keto-methylene pseudopeptides (Ewenson et al. (1986) *J Med Chem* 29:295; and Ewenson et al. in *Peptides: Structure and Function* (Proceedings of the 9th American Peptide Symposium) Pierce Chemical Co. Rockland, IL, 1985),  $\beta$ -turn dipeptide cores (Nagai et al. (1985) *Tetrahedron Lett* 26:647; and Sato et al. (1986) *J Chem Soc Perkin Trans* 1:1231), and  $\beta$ -aminoalcohols (Gordon et al. (1985) *Biochem Biophys Res Commun* 126:419; and Dann et al. (1986) *Biochem Biophys Res Commun* 134:71).

#### VI. Vaccine Formulations for *H. pylori* Nucleic Acids and Polypeptides

This invention also features vaccine compositions for protection against infection by *H. pylori* or for treatment of *H. pylori* infection, a gram-negative spiral microaerophilic bacterium. In one embodiment, the vaccine compositions contain one or more immunogenic components such as a surface protein from *H. pylori*, or portion thereof, and a pharmaceutically acceptable carrier. Nucleic acids within the scope of the invention are exemplified by the nucleic acids shown in the Sequence Listing which encode *H. pylori* surface proteins. However, any nucleic acid encoding an immunogenic *H. pylori* protein, or portion thereof, which is capable of expression in a cell, can be used in the present invention. These vaccines have therapeutic and prophylactic utilities.

One aspect of the invention provides a vaccine composition for protection against infection by *H. pylori* which contains at least one immunogenic fragment of an *H. pylori* protein and a pharmaceutically acceptable carrier. Preferred fragments include peptides of at least about 10 amino acid residues in length, preferably about 10-20 amino acid residues in length, and more preferably about 12-16 amino acid residues in length.

Immunogenic components of the invention can be obtained, for example, by screening polypeptides recombinantly produced from the corresponding fragment of the nucleic acid encoding the full-length *H. pylori* protein. In addition, fragments can be chemically synthesized using techniques known in the art such as conventional Merrifield solid phase f-Moc or t-Boc chemistry.

In one embodiment, immunogenic components are identified by the ability of the peptide to stimulate T cells. Peptides which stimulate T cells, as determined by, for example, T cell proliferation or cytokine secretion are defined herein as comprising at least one T cell epitope. T cell epitopes are believed to be involved in initiation and

- 70 -

perpetuation of the immune response to the protein allergen which is responsible for the clinical symptoms of allergy. These T cell epitopes are thought to trigger early events at the level of the T helper cell by binding to an appropriate HLA molecule on the surface of an antigen presenting cell, thereby stimulating the T cell subpopulation with the relevant T cell receptor for the epitope. These events lead to T cell proliferation, lymphokine secretion, local inflammatory reactions, recruitment of additional immune cells to the site of antigen/T cell interaction, and activation of the B cell cascade, leading to the production of antibodies. A T cell epitope is the basic element, or smallest unit of recognition by a T cell receptor, where the epitope comprises amino acids essential to receptor recognition (e.g., approximately 6 or 7 amino acid residues). Amino acid sequences which mimic those of the T cell epitopes are within the scope of this invention.

In another embodiment, immunogenic components of the invention are identified through genomic vaccination. The basic protocol is based on the idea that expression libraries consisting of all or parts of a pathogen genome, e.g., an *H. pylori* genome, can confer protection when used to genetically immunize a host. This expression library immunization (ELI) is analogous to expression cloning and involves reducing a genomic expression library of a pathogen, e.g., *H. pylori*, into plasmids that can act as genetic vaccines. The plasmids can also be designed to encode genetic adjuvants which can dramatically stimulate the humoral response. These genetic adjuvants can be introduced at remote sites and act as well extracellularly as intracellularly.

This is a new approach to vaccine production that has many of the advantages of live/attenuated pathogens but no risk of infection. An expression library of pathogen DNA is used to immunize a host thereby producing the effects of antigen presentation of a live vaccine without the risk. For example, in the present invention, random fragments from the *H. pylori* genome or from cosmid or plasmid clones, as well as PCR products from genes identified by genomic sequencing, can be used to immunize a host. The feasibility of this approach has been demonstrated with *Mycoplasma pulmonis* (Barry et al., *Nature* 377:632-635, 1995), where even partial expression libraries of *Mycoplasma pulmonis*, a natural pathogen in rodents, provided protection against challenge from the pathogen.

ELI is a technique that allows for production of a non-infectious multipartite vaccine, even when little is known about pathogen's biology, because ELI uses the immune system to screen candidate genes. Once isolated, these genes can be used as genetic vaccines or for development of recombinant protein vaccines. Thus, ELI allows for production of vaccines in a systematic, largely mechanized fashion.

- 71 -

Screening immunogenic components can be accomplished using one or more of several different assays. For example, *in vitro*, peptide T cell stimulatory activity is assayed by contacting a peptide known or suspected of being immunogenic with an antigen presenting cell which presents appropriate MHC molecules in a T cell culture.

5 Presentation of an immunogenic *H. pylori* peptide in association with appropriate MHC molecules to T cells in conjunction with the necessary costimulation has the effect of transmitting a signal to the T cell that induces the production of increased levels of cytokines, particularly of interleukin-2 and interleukin-4. The culture supernatant can be obtained and assayed for interleukin-2 or other known cytokines. For example, any one  
10 of several conventional assays for interleukin-2 can be employed, such as the assay described in *Proc. Natl. Acad. Sci USA*, 86: 1333 (1989) the pertinent portions of which are incorporated herein by reference. A kit for an assay for the production of interferon is also available from Genzyme Corporation (Cambridge, MA).

Alternatively, a common assay for T cell proliferation entails measuring tritiated  
15 thymidine incorporation. The proliferation of T cells can be measured *in vitro* by determining the amount of <sup>3</sup>H-labeled thymidine incorporated into the replicating DNA of cultured cells. Therefore, the rate of DNA synthesis and, in turn, the rate of cell division can be quantified.

Vaccine compositions of the invention containing immunogenic components  
20 (e.g., *H. pylori* polypeptide or fragment thereof or nucleic acid encoding an *H. pylori* polypeptide or fragment thereof) preferably include a pharmaceutically acceptable carrier. The term "pharmaceutically acceptable carrier" refers to a carrier that does not cause an allergic reaction or other untoward effect in patients to whom it is administered. Suitable pharmaceutically acceptable carriers include, for example, one or more of  
25 water, saline, phosphate buffered saline, dextrose, glycerol, ethanol and the like, as well as combinations thereof. Pharmaceutically acceptable carriers may further comprise minor amounts of auxiliary substances such as wetting or emulsifying agents, preservatives or buffers, which enhance the shelf life or effectiveness of the antibody. For vaccines of the invention containing *H. pylori* polypeptides, the polypeptide is  
30 coadministered with a suitable adjuvant.

It will be apparent to those of skill in the art that the therapeutically effective amount of DNA or protein of this invention will depend, *inter alia*, upon the administration schedule, the unit dose of antibody administered, whether the protein or DNA is administered in combination with other therapeutic agents, the immune status  
35 and health of the patient, and the therapeutic activity of the particular protein or DNA.

- 72 -

Vaccine compositions are conventionally administered parenterally, e.g., by injection, either subcutaneously or intramuscularly. Methods for intramuscular immunization are described by Wolff et al. (1990) *Science* 247: 1465-1468 and by Sedegah et al. (1994) *Immunology* 91: 9866-9870. Other modes of administration include oral and pulmonary formulations, suppositories, and transdermal applications. Oral immunization is preferred over parenteral methods for inducing protection against infection by *H. pylori*. Czinn et. al. (1993) *Vaccine* 11: 637-642. Oral formulations include such normally employed excipients as, for example, pharmaceutical grades of mannitol, lactose, starch, magnesium stearate, sodium saccharine, cellulose, magnesium carbonate, and the like.

The vaccine compositions of the invention can include an adjuvant, including, but not limited to aluminum hydroxide; N-acetyl-muramyl--L-threonyl-D-isoglutamine (thr-MDP); N-acetyl-nor-muramyl-L-alanyl-D-isoglutamine (CGP 11637, referred to as nor-MDP); N-acetylmuramyl-L-alanyl-D-isoglutaminyl-L-alanine-2-(1'-2'-dipalmitoyl-sn-glycero-3-hydroxyphosphoryloxy)-ethylamine (CGP 19835A, referred to as MTP-PE); RIBI, which contains three components from bacteria; monophosphoryl lipid A; trehalose dimycolate; cell wall skeleton (MPL + TDM + CWS) in a 2% squalene/Tween 80 emulsion; and cholera toxin. Others which may be used are non-toxic derivatives of cholera toxin, including its B subunit, and/or conjugates or genetically engineered fusions of the *H. pylori* polypeptide with cholera toxin or its B subunit, procholeraenoid, fungal polysaccharides, including schizophyllan, muramyl dipeptide, muramyl dipeptide derivatives, phorbol esters, labile toxin of *E. coli*, non-*H. pylori* bacterial lysates, block polymers or saponins.

Other suitable delivery methods include biodegradable microcapsules or immuno-stimulating complexes (ISCOMs) or liposomes, genetically engineered attenuated live vectors such as viruses or bacteria, and recombinant (chimeric) virus-like particles, e.g., bluetongue. The amount of adjuvant employed will depend on the type of adjuvant used. For example, when the mucosal adjuvant is cholera toxin, it is suitably used in an amount of 5 µg to 50 µg, for example 10 µg to 35 µg. When used in the form of microcapsules, the amount used will depend on the amount employed in the matrix of the microcapsule to achieve the desired dosage. The determination of this amount is within the skill of a person of ordinary skill in the art.

Carrier systems in humans may include enteric release capsules protecting the antigen from the acidic environment of the stomach, and including *H. pylori* polypeptide in an insoluble form as fusion proteins. Suitable carriers for the vaccines of the

- 73 -

invention are enteric coated capsules and polylactide-glycolide microspheres. Suitable diluents are 0.2 N NaHCO<sub>3</sub> and/or saline.

Vaccines of the invention can be administered as a primary prophylactic agent in adults or in children, as a secondary prevention, after successful eradication of *H. pylori* in an infected host, or as a therapeutic agent in the aim to induce an immune response in a susceptible host to prevent infection by *H. pylori*. The vaccines of the invention are administered in amounts readily determined by persons of ordinary skill in the art. Thus, for adults a suitable dosage will be in the range of 10 µg to 10 g, preferably 10 µg to 100 mg, for example 50 µg to 50 mg. A suitable dosage for adults will also be in the range of 5 µg to 500 mg. Similar dosage ranges will be applicable for children. Those skilled in the art will recognize that the optimal dose may be more or less depending upon the patient's body weight, disease, the route of administration, and other factors. Those skilled in the art will also recognize that appropriate dosage levels can be obtained based on results with known oral vaccines such as, for example, a vaccine based on an *E. coli* lysate (6 mg dose daily up to total of 540 mg) and with an enterotoxigenic *E. coli* purified antigen (4 doses of 1 mg) (Schulman et al., *J. Urol.* 150:917-921 (1993); Boedecker et al., *American Gastroenterological Assoc.* 999:A-222 (1993)). The number of doses will depend upon the disease, the formulation, and efficacy data from clinical trials. Without intending any limitation as to the course of treatment, the treatment can be administered over 3 to 8 doses for a primary immunization schedule over 1 month (Boedeker, *American Gastroenterological Assoc.* 888:A-222 (1993)).

It will be apparent to those skilled in the art that some of the vaccine compositions of the invention are useful only for preventing *H. pylori* infection, some are useful only for treating *H. pylori* infection, and some are useful for both preventing and treating *H. pylori* infection. In a preferred embodiment, the vaccine composition of the invention provides protection against *H. pylori* infection by stimulating humoral and/or cell-mediated immunity against *H. pylori*. It should be understood that amelioration of any of the symptoms of *H. pylori* infection is a desirable clinical goal, including a lessening of the dosage of medication used to treat *H. pylori*-caused disease.

#### VII. Antibodies Reactive With *H. pylori* Polypeptides

The invention also includes antibodies specifically reactive with the subject *H. pylori* polypeptide. Anti-protein/anti-peptide antisera or monoclonal antibodies can be made by standard protocols (See, for example, *Antibodies: A Laboratory Manual* ed. by Harlow and Lane (Cold Spring Harbor Press: 1988)). A mammal such as a mouse, a

- 74 -

hamster or rabbit can be immunized with an immunogenic form of the peptide. Techniques for conferring immunogenicity on a protein or peptide include conjugation to carriers or other techniques well known in the art. An immunogenic portion of the subject *H. pylori* polypeptide can be administered in the presence of adjuvant. The progress of immunization can be monitored by detection of antibody titers in plasma or serum. Standard ELISA or other immunoassays can be used with the immunogen as antigen to assess the levels of antibodies.

In a preferred embodiment, the subject antibodies are immunospecific for antigenic determinants of the *H. pylori* polypeptides of the invention, e.g. antigenic determinants of a polypeptide shown in the Sequence Listing, or a closely related human or non-human mammalian homolog (e.g., 90% homologous, more preferably at least 95% homologous). In yet a further preferred embodiment of the invention, the anti-*H. pylori* antibodies do not substantially cross react (i.e., react specifically) with a protein which is for example, less than 80% percent homologous to a sequence shown in the Sequence Listing. By "not substantially cross react", it is meant that the antibody has a binding affinity for a non-homologous protein which is less than 10 percent, more preferably less than 5 percent, and even more preferably less than 1 percent, of the binding affinity for a protein contained in the Sequence Listing. In a most preferred embodiment, there is no crossreactivity between bacterial and mammalian antigens.

The term antibody as used herein is intended to include fragments thereof which are also specifically reactive with *H. pylori* polypeptides. Antibodies can be fragmented using conventional techniques and the fragments screened for utility in the same manner as described above for whole antibodies. For example, F(ab')<sub>2</sub> fragments can be generated by treating antibody with pepsin. The resulting F(ab')<sub>2</sub> fragment can be treated to reduce disulfide bridges to produce Fab' fragments. The antibody of the invention is further intended to include bispecific and chimeric molecules having an anti-*H. pylori* portion.

Both monoclonal and polyclonal antibodies (Ab) directed against *H. pylori* polypeptides or *H. pylori* polypeptide variants, and antibody fragments such as Fab' and F(ab')<sub>2</sub>, can be used to block the action of *H. pylori* polypeptide and allow the study of the role of a particular *H. pylori* polypeptide of the invention in aberrant or unwanted intracellular signaling, as well as the normal cellular function of the *H. pylori* and by microinjection of anti-*H. pylori* polypeptide antibodies of the present invention.

Antibodies which specifically bind *H. pylori* epitopes can also be used in immunohistochemical staining of tissue samples in order to evaluate the abundance and pattern of expression of *H. pylori* antigens. Anti *H. pylori* polypeptide antibodies can be

- 75 -

used diagnostically in immuno-precipitation and immuno-blotting to detect and evaluate *H. pylori* levels in tissue or bodily fluid as part of a clinical testing procedure. Likewise, the ability to monitor *H. pylori* polypeptide levels in an individual can allow determination of the efficacy of a given treatment regimen for an individual afflicted with such a disorder. The level of an *H. pylori* polypeptide can be measured in cells found in bodily fluid, such as in urine samples or can be measured in tissue, such as produced by gastric biopsy. Diagnostic assays using anti-*H. pylori* antibodies can include, for example, immunoassays designed to aid in early diagnosis of *H. pylori* infections. The present invention can also be used as a method of detecting antibodies contained in samples from individuals infected by this bacterium using specific *H. pylori* antigens.

Another application of anti-*H. pylori* polypeptide antibodies of the invention is in the immunological screening of cDNA libraries constructed in expression vectors such as  $\lambda$ gt11,  $\lambda$ gt18-23,  $\lambda$ ZAP, and  $\lambda$ ORF8. Messenger libraries of this type, having coding sequences inserted in the correct reading frame and orientation, can produce fusion proteins. For instance,  $\lambda$ gt11 will produce fusion proteins whose amino termini consist of  $\beta$ -galactosidase amino acid sequences and whose carboxy termini consist of a foreign polypeptide. Antigenic epitopes of a subject *H. pylori* polypeptide can then be detected with antibodies, as, for example, reacting nitrocellulose filters lifted from infected plates with anti-*H. pylori* polypeptide antibodies. Phage, scored by this assay, can then be isolated from the infected plate. Thus, the presence of *H. pylori* gene homologs can be detected and cloned from other species, and alternate isoforms (including splicing variants) can be detected and cloned.

#### VIII. Kits Containing Nucleic Acids, Polypeptides or Antibodies of the Invention

The nucleic acid, polypeptides and antibodies of the invention can be combined with other reagents and articles to form kits. Kits for diagnostic purposes typically comprise the nucleic acid, polypeptides or antibodies in vials or other suitable vessels. Kits typically comprise other reagents for performing hybridization reactions, polymerase chain reactions (PCR), or for reconstitution of lyophilized components, such as aqueous media, salts, buffers, and the like. Kits may also comprise reagents for sample processing such as detergents, chaotropic salts and the like. Kits may also comprise immobilization means such as particles, supports, wells, dipsticks and the like. Kits may also comprise labeling means such as dyes, developing reagents, radioisotopes, fluorescent agents, luminescent or chemiluminescent agents, enzymes, intercalating agents and the like. With the nucleic acid and amino acid sequence information provided

- 76 -

herein, individuals skilled in art can readily assemble kits to serve their particular purpose. Kits further can include instructions for use.

#### IX. Drug Screening Assays Using *H. pylori* Polypeptides

5 By making available purified and recombinant *H. pylori* polypeptides, the present invention provides assays which can be used to screen for drugs which are either agonists or antagonists of the normal cellular function, in this case, of the subject *H. pylori* polypeptides, or of their role in intracellular signaling. Such inhibitors or potentiators may be useful as new therapeutic agents to combat *H. pylori* infections in  
10 humans. A variety of assay formats will suffice and, in light of the present inventions, will be comprehended by the skilled artisan.

In many drug screening programs which test libraries of compounds and natural extracts, high throughput assays are desirable in order to maximize the number of compounds surveyed in a given period of time. Assays which are performed in cell-free  
15 systems, such as may be derived with purified or semi-purified proteins, are often preferred as "primary" screens in that they can be generated to permit rapid development and relatively easy detection of an alteration in a molecular target which is mediated by a test compound. Moreover, the effects of cellular toxicity and/or bioavailability of the test compound can be generally ignored in the *in vitro* system, the assay instead being  
20 focused primarily on the effect of the drug on the molecular target as may be manifest in an alteration of binding affinity with other proteins or change in enzymatic properties of the molecular target. Accordingly, in an exemplary screening assay of the present invention, the compound of interest is contacted with an isolated and purified *H. pylori* polypeptide.

25 Screening assays can be constructed *in vitro* with a purified *H. pylori* polypeptide or fragment thereof, such as an *H. pylori* polypeptide having enzymatic activity, such that the activity of the polypeptide produces a detectable reaction product. The efficacy of the compound can be assessed by generating dose response curves from data obtained using various concentrations of the test compound. Moreover, a control  
30 assay can also be performed to provide a baseline for comparison. Suitable products include those with distinctive absorption, fluorescence, or chemi-luminescence properties, for example, because detection may be easily automated. A variety of synthetic or naturally occurring compounds can be tested in the assay to identify those which inhibit or potentiate the activity of the *H. pylori* polypeptide. Some of these  
35 active compounds may directly, or with chemical alterations to promote membrane

- 77 -

permeability or solubility, also inhibit or potentiate the same activity (e.g., enzymatic activity) in whole, live *H. pylori* cells.

This invention is further illustrated by the following examples which should not be construed as limiting. The contents of all references and published patent applications cited throughout this application are hereby incorporated by reference.

### Other Embodiments

Many of the nucleic acids and corresponding polypeptides of the invention were disclosed previously in the parent applications, U.S.S.N. 08/761,318, filed December 6, 1996 (Attorney Docket No.: GTN-009CP2), U.S.S.N. 08/736,905, filed October 25, 1996 (Attorney Docket No.: GTN-010CP) and U.S.S.N. 08/738,859, filed October 28, 1996 (Attorney Docket No.: GTN-009CP), which are a continuation-in-part of U.S.S.N. 08/625,811, filed March 29, 1996 (Attorney Docket No.: GTN-009), and U.S.S.N. 08/758,731, filed April 2, 1996 (Attorney Docket No.: GTN-010). The correlation between sequence identification numbers in the above-identified parent applications and sequence identification numbers provided herein is outlined in Table 7 below.

TABLE 7

ORF Name	Source	Parent nt Seq ID#	Parent aa Seq ID#	CIP1 nt Seq ID#	CIP1 aa Seq ID#	Current nt Seq ID#	Current aa Seq ID#
01ael2021orf5	009CP1	5	867	3	1095	1	492
01cel0516orf15	009CP1	15	877	10	1102	2	493
01cel0516orf19	009CP1	18	880	13	1105	3	494
01cel0516orf20	009CP1	19	881	14	1106	4	495
01cel0516orf21	009CP1	20	882	15	1107	5	496
01cel0516orf24	009CP1	21	883	16	1108	6	497
01cel1618orf24	009CP1	36	898	30	1122	7	498
01cel1618orf3	009CP1	37	899	31	1123	8	499
01cel1513orf17	010CP1	44	1957	35	1554	9	500
01cp11414orf2	009CP1	44	906	36	1128	10	501
01cp11710orf1	009CP1	46	908	37	1129	11	502
01cel1618orf1	010CP1	51	1964	38	1557	12	503
01cel1618orf13	010CP1	54	1967	40	1559	13	504
01ep11504orf5	009CP1	65	927	52	1144	14	505
01ep30520orf24	009CP1	67	929	54	1146	15	506
01cp11108orf6	010CP1	74	1987	54	1573	16	507
01gpl1016orf13	009CP1	78	940	63	1155	17	508
02ael1612orf1	009CP1	97	959	79	1171	18	509
02ael1612orf15	009CP1	99	961	80	1172	19	510
01ep30520orf8	010CP1	121	2034	86	1605	20	511
02ap21113orf2	009CP1	112	974	88	1180	21	512
02ap71220orf2	009CP1	114	976	90	1182	22	513
01gel0801orf2	010CP1	133	2046	91	1610	23	514

- 78 -

02ce10213orf19	009CP1	116	978	92	1184	24	515
02ce10216orf6	009CP1	123	985	96	1188	25	516
02ce10216orf7	009CP1	124	986	97	1189	26	517
02ce10809orf6	009CP1	130	992	103	1195	27	518
04ce11617orf27	001CP7	1087	1538	119	372	28	519
02ep30607orf27	009CP1	158	1020	120	1212	29	520
02ge20116orf25	009CP1	164	1026	125	1217	30	521
02ae21214orf1	010CP1	193	2106	129	1648	31	522
03ap21820orf4	009CP1	181	1043	139	1231	32	523
03ap21820orf8	009CP1	183	1045	141	1233	33	524
02ce10114orf3	010CP1	206	2119	142	1661	34	525
03ce21717orf1	009CP1	186	1048	144	1236	35	526
03gp20123orf3	009CP1	196	1058	152	1244	36	527
02ce10216orf4	010CP1	225	2138	152	1671	37	528
03xe11215orf7	009CP1	199	1061	154	1246	38	529
02ce11022orf2	010CP1	232	2145	159	1678	39	530
04ce11617orf16	009CP1	207	1069	161	1253	40	531
04ce11617orf26	009CP1	211	1073	162	1254	41	532
14ce21516orf1	001CP7	1306	1757	163	416	42	533
04ce11617orf4	009CP1	214	1076	164	1256	43	534
04ee70114orf10	001CP7	220	1082	169	1261	44	535
hp4e14535orf3	009CP1	1414	1865	181	434	45	536
04ge11713orf9	009CP1	246	1108	185	1277	46	537
04gp11213orf11	009CP1	247	1109	186	1278	47	538
04gp11213orf14	009CP1	248	1110	187	1279	48	539
04gp11213orf22	009CP1	249	1111	188	1280	49	540
04gp11803orf13	009CP1	254	1116	191	1283	50	541
05ae20220orf119	009CP1	263	1125	195	1287	51	542
05ae20220orf56	009CP1	270	1132	197	1289	52	543
05ae20220orf58	009CP1	271	1133	198	1290	53	544
05ae20220orf95	010CP1	280	1142	201	1293	54	545
02ep30607orf32	009CP1	297	2210	204	1723	55	546
05ap21216orf4	009CP1	290	1152	206	1298	56	547
05ce10420orf1	009CP1	291	1153	207	1299	57	548
05cp11911orf35	009CP1	297	1159	209	1301	58	549
05cp20518orf39	009CP1	305	1167	212	1304	59	550
05ee10411orf5	009CP1	308	1170	213	1305	60	551
05ep11717orf2	009CP1	314	1176	219	1311	61	552
05gp11901orf19	009CP1	324	1186	225	1317	62	553
05gp20111orf12	009CP1	326	1188	227	1319	63	554
05gp20111orf4	009CP1	328	1190	229	1321	64	555
05gp20111orf6	009CP1	329	1191	230	1322	65	556
06ce10808orf2	009CP1	343	1205	243	1335	66	557
06cp11118orf7	009CP1	352	1214	251	1343	67	558
06cp30603orf15	009CP1	364	1226	262	1354	68	559
06cp30603orf16	009CP1	365	1227	263	1355	69	560
06ep10306orf5	009CP1	373	1235	266	1358	70	561
06ep11108orf17	009CP1	378	1240	270	1362	71	562
06ge10115orf17	009CP1	386	1248	277	1369	72	563
06ge10115orf19	009CP1	387	1249	278	1370	73	564
06ge10115orf4	009CP1	388	1250	279	1371	74	565
07ap11111orf13	009CP1	408	1270	297	1389	75	566

- 79 -

07ee11519orf1	009CPI	422	1284	304	1396	76	567
07gp11807orf17	009CPI	448	1310	323	1415	77	568
07gp31516orf9	009CPI	453	1315	326	1418	78	569
09ap20802orf14	010CPI	465	1327	337	1429	79	570
05ap10914orf3	010CPI	569	2482	345	1864	80	571
05ce10613orf1	010CPI	586	2499	349	1868	81	572
05ce10613orf2	009CPI	587	2500	350	1869	82	573
09ge70821orf2	009CPI	491	1353	354	1446	83	574
11ap11902orf3	009CPI	497	1359	357	1449	84	575
11cp12006orf13	009CPI	521	1383	381	1473	85	576
11ee10423orf1	009CPI	530	1392	390	1482	86	577
11ee10423orf4	009CPI	531	1393	391	1483	87	578
11ge10309orf12	009CPI	534	1396	394	1486	88	579
11ge10309orf7	009CPI	551	1413	395	1487	89	580
12ap10324orf6	009CPI	565	1427	398	1490	90	581
12ge10305orf5	009CPI	573	1435	403	1495	91	582
13ae10511orf3	009CPI	587	1449	410	1502	92	583
13ae10712orf15	009CPI	592	1454	413	1505	93	584
13ap11517orf15	009CPI	599	1461	418	1510	94	585
13ee10216orf7	009CPI	610	1472	422	1514	95	586
13ep12003orf21	009CPI	625	1487	435	1527	96	587
14ae21813orf2	009CPI	630	1492	438	1530	97	588
14ce10720orf12	001CP6	631	1493	439	1531	98	589
02ep30607orf19	009CPI	1065	1516	440	1372	99	590
14ce10720orf3	009CPI	632	1494	440	1532	100	591
14cp10705orf4	009CPI	645	1507	447	1539	101	592
14ee10308orf1	009CPI	649	1511	451	1543	102	593
14ee10308orf4	010CPI	651	1513	452	1544	103	594
06cp11722orf11	009CPI	739	2652	454	1973	104	595
14ee11217orf2	009CPI	655	1517	456	1548	105	596
14ee11217orf3	010CPI	656	1518	457	1549	106	597
06cp11722orf5	009CPI	745	2658	460	1979	107	598
14ge10705orf14	009CPI	660	1522	461	1553	108	599
14gp11820orf1	001CP6	663	1525	464	1556	109	600
04ge11713orf35	010CPI	1098	1549	466	1398	110	601
06cp20302orf3	010CPI	751	2664	466	1985	111	602
06cp20302orf5	010CPI	752	2665	467	1986	112	603
06cp20302orf6	010CPI	753	2666	468	1987	113	604
06cp20302orf7	010CPI	754	2667	469	1988	114	605
06ee10207orf1	009CPI	763	2676	471	1990	115	606
29ge10307orf3	010CPI	679	1541	475	1567	116	607
06ee11611orf1	009CPI	770	2683	475	1994	117	608
29ge10307orf4	010CPI	680	1542	476	1568	118	609
06ep10306orf10	010CPI	773	2686	477	1996	119	610
06ep10306orf11	009CPI	774	2687	478	1997	120	611
hpl1e10523orf3	009CPI	687	1549	481	1573	121	612
hpl1p13852orf4	009CPI	700	1562	494	1586	122	613
hpl1p13852orf6	009CPI	701	1563	495	1587	123	614
hpl1p13852orf7	009CPI	702	1564	496	1588	124	615
hpl1p13947orf11	009CPI	708	1570	502	1594	125	616
hpl1p14013orf17	009CPI	715	1577	508	1600	126	617
hp2e10229orf2	009CPI	716	1578	509	1601	127	618

- 80 -

hp2e10229orf4	009CP1	717	1579	510	1602	128	619
hp2e10911orf24	010CP1	722	1584	515	1607	129	620
06gp10108orf2	001CP6	816	2729	516	2035	130	621
07ap11015orf2	009CP1	1167	1618	520	1452	131	622
hp2e11858orf6	001CP6	730	1592	520	1612	132	623
07ap11015orf4	009CP1	1168	1619	521	1453	133	624
hp2p10272orf1	009CP1	733	1595	522	1614	134	625
hp2p10625orf5	009CP1	742	1604	525	1617	135	626
hp2p10625orf6	009CP1	743	1605	526	1618	136	627
hp3e10349orf27	009CP1	751	1613	530	1622	137	628
hp3e11024orf34	009CP1	763	1625	539	1631	138	629
hp3e11024orf49	009CP1	767	1629	543	1635	139	630
hp3e11024orf5	009CP1	768	1630	544	1636	140	631
hp3e11075orf3	010CP1	773	1635	546	1638	141	632
07ap80601orf2	010CP1	863	2776	548	2067	142	633
07ce10203orf11	010CP1	871	2784	554	2073	143	634
07ce10203orf17	009CP1	873	2786	556	2075	144	635
hp3p11086orf1	010CP1	796	1658	561	1653	145	636
07ce11206orf1	009CP1	884	2797	567	2086	146	637
hp4p13446orf13	009CP1	808	1670	571	1663	147	638
11ce10917orf10	009CP1	498	1360	572	1504	148	639
hp4p13446orf3	009CP1	809	1671	572	1664	149	640
hp5e15211orf23	010CP1	827	1689	581	1673	150	641
07cp11213orf1	009CP1	905	2818	584	2103	151	642
hp5p15612orf2	009CP1	848	1710	594	1686	152	643
hp5p15641orf23	001CP6	851	1713	596	1688	153	644
12ap10324orf5	010CP1	1265	1716	598	1530	154	645
07ep30818orf4	009CP1	932	2845	600	2119	155	646
01ce11618orf11	009CP1	52	1965	607	1699	156	647
01ce11618orf22	009CP1	61	1974	608	1700	157	648
01ce11618orf27	009CP1	62	1975	609	1701	158	649
01ce11618orf7	009CP1	67	1980	610	1702	159	650
01ce11618orf9	009CP1	68	1981	611	1703	160	651
01cp11710orf16	009CP1	84	1997	612	1704	161	652
01cp11710orf18	009CP1	86	1999	613	1705	162	653
01cp11710orf5	009CP1	90	2003	615	1707	163	654
01cp11710orf6	009CP1	91	2004	616	1708	164	655
01cp11710orf9	009CP1	93	2006	617	1709	165	656
01ge10203orf14	009CP1	125	2038	618	1710	166	657
01ge10203orf6	009CP1	128	2041	619	1711	167	658
01ge10203orf7	009CP1	129	2042	620	1712	168	659
03ge10505orf2	009CP1	376	2289	644	1736	169	660
04ce11617orf10	001CP6	404	2317	645	1737	170	661
14gp12015orf13	001CP6	1330	1781	648	1580	171	662
14gp12015orf16	009CP1	1332	1783	650	1582	172	663
04ge11210orf1	009CP1	449	2362	651	1743	173	664
04ge11713orf10	009CP1	455	2368	653	1745	174	665
04ge11713orf28	009CP1	462	2375	655	1747	175	666
04ge11713orf36	010CP1	464	2377	656	1748	176	667
09ap11406orf2	009CP1	1014	2927	657	2176	177	668
04gp11213orf5	001CP6	487	2400	661	1753	178	669
29ge10111orf1	009CP1	1353	1804	669	1601	179	670

- 81 -

05ae20220orf37	001CP6	548	2461	675	1767	180	671
hp2e10911orf25	009CP1	1377	1828	686	1618	181	672
06ap10209orf1	009CP1	697	2610	690	1782	182	673
06ap10209orf4	009CP1	699	2612	691	1783	183	674
06ee10709orf2	009CP1	767	2680	695	1787	184	675
06ep10306orf1	009CP1	772	2685	696	1788	185	676
07cp10312orf5	009CP1	898	2811	703	1795	186	677
07cp21714orf1	009CP1	908	2821	706	1798	187	678
07cp21714orf3	009CP1	910	2823	708	1800	188	679
09ap20802orf13	009CP1	1025	2938	713	1805	189	680
09ap20802orf21	010CP1	1031	2944	714	1806	190	681
11ce10908orf1	001CP6	1100	3013	716	2235	191	682
hp4p11352orf9	009CP1	1419	1870	718	1650	192	683
11ge10309orf24	009CP1	1155	3068	725	1817	193	684
11ge10309orf28	009CP1	1157	3070	727	1819	194	685
11ge10309orf39	009CP1	1164	3077	731	1823	195	686
11ge10309orf4	009CP1	1165	3078	732	1824	196	687
11ge10309orf5	010CP1	1166	3079	733	1825	197	688
11cp12002orf3	009CP1	1129	3042	745	2264	198	689
13ae10610orf1	010CP1	1245	3158	747	1839	199	690
11ee10423orf2	009CP1	1137	3050	753	2272	200	691
14cp10119orf1	010CP1	1349	3262	758	1850	201	692
11ge10308orf1	009CP1	1145	3058	761	2280	202	693
14ee10308orf7	009CP1	1375	3288	762	1854	203	694
14ge10705orf3	009CP1	1411	3324	763	1855	204	695
hp1e10506orf5	009CP1	1495	3408	766	1858	205	696
hp3p10807orf4	009CP1	1746	3659	775	1867	206	697
hp3p10807orf7	009CP1	1749	3662	776	1868	207	698
hp4e12063orf1	010CP1	1754	3667	777	1869	208	699
12ap10324orf8	009CP1	1200	3113	778	2297	209	700
hp4p11352orf5	010CP1	1783	3696	782	1874	210	701
12ge10610orf2	010CP1	1217	3130	784	2303	211	702
13ee12016orf7	010CP1	1312	3225	844	2363	212	703
14ee10419orf5	010CP1	1384	3297	898	2417	213	704
14ee11217orf4	010CP1	1390	3303	904	2423	214	705
14ee21118orf1	010CP1	1391	3304	905	2424	215	706
14ee21118orf2	010CP1	1392	3305	906	2425	216	707
14ep11115orf3	010CP1	1397	3310	911	2430	217	708
14ep11905orf13	010CP1	1400	3313	914	2433	218	709
14ep11905orf9	010CP1	1408	3321	922	2441	219	710
29ap10306orf3	010CP1	1462	3375	957	2476	220	711
29ap11902orf1	010CP1	1463	3376	958	2477	221	712
29ep20112orf2	010CP1	1469	3382	964	2483	222	713
29ge10111orf3	010CP1	1471	3384	966	2485	223	714
hp1e10554orf1	010CP1	1500	3413	982	2501	224	715
hp1p10543orf4	010CP1	1506	3419	988	2507	225	716
hp1p13947orf10	010CP1	1552	3465	1025	2544	226	717
hp2e10911orf35	010CP1	1592	3505	1047	2566	227	718
hp3e10057orf3	010CP1	1617	3530	1062	2581	228	719
hp3e10128orf1	010CP1	1619	3532	1064	2583	229	720
hp3e10302orf17	010CP1	1627	3540	1072	2591	230	721
hp3e10302orf26	010CP1	1633	3546	1078	2597	231	722

- 82 -

hp3e11122orf3	010CPI	1693	3606	1127	2646	232	723
hp3p10156orf2	010CPI	1715	3628	1141	2660	233	724
hp3p10156orf3	010CPI	1716	3629	1142	2661	234	725
hp3p10156orf7	010CPI	1718	3631	1144	2663	235	726
hp3p10304orf2	010CPI	1724	3637	1148	2667	236	727
hp3p10807orf6	010CPI	1748	3661	1160	2679	237	728
hp3p11086orf2	010CPI	1753	3666	1164	2683	238	729
hp4e14535orf7	010CPI	1776	3689	1178	2697	239	730
hp4e14535orf8	010CPI	1777	3690	1179	2698	240	731
hp4p11393orf1	010CPI	1785	3698	1182	2701	241	732
hp4p11393orf2	010CPI	1786	3699	1183	2702	242	733
hp4p11393orf6	010CPI	1790	3703	1187	2706	243	734
hp4p12005orf2	010CPI	1792	3705	1189	2708	244	735
hp4p13446orf5	010CPI	1806	3719	1203	2722	245	736
hp5p15580orf1	010CPI	1889	3802	1258	2777	246	737
hp5p15653orf1	010CPI	1898	3811	1264	2783	247	738
hp5p15653orf2	010CPI	1899	3812	1265	2784	248	739
04ap20904orf3	010CPI	205	1067	1281	2800	249	740
06gp11920orf11	010CPI	398	1260	1288	2807	250	741
12ge20305orf26	010CPI	581	1443	1301	2820	251	742
13ee10216orf57	009CPI	609	1471	1303	2822	252	743
04cp11202_24256567_c3_117	009CPI			783	1715	253	744
01ae11010_40688_c2_38	009CPI			790	1882	254	745
01ae12001_24218781_f2_18	009CPI			791	1883	255	746
01ge11619_23711062_c3_14	009CPI			795	1887	256	747
01gp11016_4103403_c2_13	009CPI			799	1891	257	748
02ae11612_1074212_f1_1	009CPI			800	1892	258	749
02ae11612_22477267_f2_27	009CPI			802	1894	259	750
02ae11612_23598175_f1_2	009CPI			803	1895	260	751
02ae11612_33203250_c1_51	009CPI			804	1896	261	752
02ge10116_36367936_c1_19	009CPI			807	1899	262	753
02ge11622_875260_f3_36	009CPI			809	1901	263	754
02gp20706_1203402_c3_58	009CPI			810	1902	264	755
02gp20706_15781452_c2_51	009CPI			813	1905	265	756
02gp20706_23632775_f3_32	009CPI			815	1907	266	757
02gp20706_23866562_c2_53	009CPI			816	1908	267	758
02gp20706_4892558_f3_19	009CPI			821	1913	268	759
02gp20814_24415958_f3_9	009CPI			824	1916	269	760
04cp11202_24261588_f2_23	010CPI			827	1919	270	761
07ap11015_23938312_f3_2	009CPI			831	1763	271	762
05ae30220_21619067_f3_56	009CPI			835	1927	272	763
07ap11213_35401528_c1_21	009CPI			835	1767	273	764
05ae30220_24882812_c3_103	009CPI			837	1929	274	765
05ae30220_25953163_c3_98	009CPI			839	1931	275	766
05ee10816_14649077_f3_18	009CPI			841	1933	276	767
05ee10816_4103408_f2_11	009CPI			843	1935	277	768
05ee10816_4687651_c1_22	009CPI			844	1936	278	769
06ap11119_16594193_f1_9	009CPI			845	1937	279	770
06ap20306_23437632_f3_9	009CPI			851	1943	280	771
06cp30603_23452_c3_80	009CPI			854	1946	281	772
06cp30603_23476568_c1_44	009CPI			855	1947	282	773
06cp30603_4689068_c3_79	009CPI			857	1949	283	774

- 83 -

07ap11111_234693_c3_14	009CPI			860	1952	284	775
07ap20216_7227202_f3_10	009CPI			865	1957	285	776
07ep11916_5273452_c3_31	009CPI			866	1958	286	777
09cp11003_19532625_c3_17	009CPI			869	1961	287	778
09cp20502_24001388_c1_31	009CPI			871	1963	288	779
12gp31106_3024126_f2_25	009CPI			876	1968	289	780
13ae10712_14100018_f2_12	009CPI			879	1971	290	781
13ae10712_29569208_c2_27	009CPI			880	1972	291	782
14ap10815_20585777_c1_13	009CPI			884	1976	292	783
29ge30321_34157812_f3_10	009CPI			886	1978	293	784
hp1p14013_11726503_c2_20	009CPI			888	1980	294	785
hp2p10272_23697200_f3_22	009CPI			891	1983	295	786
hp2p10272_26829136_f1_1	009CPI			894	1986	296	787
hp5e15211_819455_c2_24	009CPI			897	1989	297	788
hp5p15212_34064750_f2_9	009CPI			901	1993	298	789
hp5p15641_21698387_c2_20	009CPI			904	1996	299	790
hp6e10967_23476502_f2_6	009CPI			908	2000	300	791
hp6e10967_24882750_f2_7	009CPI			909	2001	301	792
hp6e12267_4876718_f2_23	009CPI			912	2004	302	793
hp6e20339_1190660_c2_46	009CPI			913	2005	303	794
hp6e20339_21492187_c1_40	009CPI			914	2006	304	795
hp6e20339_34024187_c1_37	009CPI			916	2008	305	796
01ge10203_35281542_c3_16	009CPI			930	2022	306	797
01ge10203_860166_f3_9	009CPI			931	2023	307	798
01ge11619_13788141_c2_11	009CPI			932	2024	308	799
01ge11619_24415880_c2_12	009CPI			934	2026	309	800
01ge11619_24417813_c1_8	009CPI			935	2027	310	801
02gp20706_16803513_f1_1	009CPI			946	2038	311	802
02gp20706_20365905_f2_8	009CPI			947	2039	312	803
02gp20814_3984818_f1_1	009CPI			949	2041	313	804
04cp11202_16603425_c2_72	009CPI			955	2047	314	805
04cp11202_19797128_f1_5	009CPI			956	2048	315	806
04cp11202_23553177_c1_75	009CPI			957	2049	316	807
04cp11202_23553177_c3_109	009CPI			958	2050	317	808
04ge10816_22086531_f2_10	009CPI			963	2055	318	809
04ge10816_33726080_c2_29	009CPI			964	2056	319	810
05ae30220_14350428_f1_9	009CPI			969	2061	320	811
05ae30220_9882767_f2_34	009CPI			978	2070	321	812
05cp21223_4725443_f3_14	009CPI			981	2073	322	813
05ee10816_259703_f2_7	009CPI			982	2074	323	814
06ap11119_24426508_f3_26	009CPI			984	2076	324	815
06ee10709_21675012_f1_2	009CPI			993	2085	325	816
07ce11019_22051291_f1_1	009CPI			1004	2096	326	817
07ep11916_5913592_f3_18	009CPI			1006	2098	327	818
09cp11003_5945252_f2_4	009CPI			1013	2105	328	819
09ze10333_22460750_f2_6	009CPI			1014	2106	329	820
12ge10321_4821082_f3_14	009CPI			1028	2120	330	821
14ap10221_13689381_c3_4	009CPI			1035	2127	331	822
29ep10720_24220926_f2_8	009CPI			1046	2138	332	823
29ep10720_24432762_c3_39	009CPI			1047	2139	333	824
29ge30321_12913562_f1_1	009CPI			1050	2142	334	825
29ge30321_135253_f2_6	009CPI			1051	2143	335	826

- 84 -

29ge30321_21673965_f2_7	009CPI			1052	2144	336	827
29ge30321_24336712_f1_5	009CPI			1053	2145	337	828
hp2p10272_22692325_f3_21	009CPI			1059	2151	338	829
hp2p10272_24406280_c1_26	009CPI			1060	2152	339	830
hp3p10807_29343768_f1_1	009CPI			1061	2153	340	831
hp3p10807_29352212_f2_5	009CPI			1062	2154	341	832
hp6e20339_24317062_c3_57	009CPI			1078	2170	342	833
hp6p10233_12273302_f1_1	009CPI			1087	2179	343	834
02ep20506_24611325_f2_6	010CPI			1341	2860	344	835
07ap11213_35156577_c1_24	010CPI			1410	2929	345	836
07ap80601_976413_f3_9	010CPI			1414	2933	346	837
hp3p10807_189075_f2_4	010CPI			1464	2983	347	838
02gel1622_21695936_c1_54	010CPI			1493	3012	348	839
12gel0321_24308513_f3_20	010CPI			1504	3023	349	840
14ce21516_85786_f1_1	010CPI			1508	3027	350	841
hp6el2267_14650278_f3_29	010CPI			1517	3036	351	842
06cp30603_23476568_c2_133.aa						352	843
06ce20610_1367157_f1_8.aa						353	844
06ep30223_4698838_f2_55						354	845
12ap10324_4805318_f2_3						355	846
14ee41924_2458267_c2_93						356	847
06ep11917_24803153_c3_24						357	848
01cel1104_36125337_c1_8						358	849
01ce21104_33203250_c3_87						359	850
02ae31010_34616666_f2_27						360	851
02ae31010_35270000_f3_33						361	852
02ae31010_36132785_f2_29						362	853
02gel0116_15781452_c1_87						363	854
02gel0116_16803513_f2_34						364	855
02gel0116_16803513_f2_34						365	856
02gel0116_36367936_c1_92						366	857
03ael0804_23485968_c3_47						367	858
05cel0910_25598277_f3_3						368	859
06ael1016_30579712_f2_21						369	860
06ce20610_29298537_c2_32						370	861
06ce20610_3913967_c3_36						371	862
06ce20610_4331338_f3_18						372	863
06cp11118_212827_c1_17						373	864
06cp11217_19720300_f3_11						374	865
06cp11217_4881263_f2_9						375	866
06cp11217_4897077_f1_6						376	867
06cp30603_21492187_f2_41						377	868
06cp30603_34024187_f1_20						378	869
06cp30603_34024187_f1_20						379	870
06ep10615_14649077_f3_52						380	871
06ep10615_9842_f3_46						381	872
06ep11202_26353438_c1_22						382	873
06ep30223_23557202_c2_130						383	874
06ep30223_34409437_f3_94						384	875
06ep30223_4698838_f2_55						385	876
06ep30223_4876077_c3_149						386	877
06ep30223_5109443_c1_109						387	878

- 85 -

06ep30223_5271902_c1_106						388	879
06gp10409_3398427_f2_12						389	880
06gp71906_24261588_c2_174						390	881
06gp71906_970325_c3_190						391	882
07ae10923_24426508_f1_1						392	883
07ae10923_24426508_f1_1						393	884
09ce10413_414011_f1_3						394	885
09ce10413_5865665_f1_4						395	886
09ce52017_29324062_c1_21						396	887
09cp10224_1062966_c3_61						397	888
09cp10224_1412715_c3_56						398	889
09cp10224_429510_c2_46.aa						399	890
09cp10224_4484718_c1_38						400	891
09cp21607_7224187_c2_12						401	892
09cp61003_14562637_c2_93						402	893
09cp61003_19532625_c1_78						403	894
09cp61003_24063587_c1_74						404	895
09cp61003_24335762_c3_111						405	896
09cp61003_5945252_f1_5						406	897
11ae80818_11188791_c3_60						407	898
11ae80818_783127_c3_63						408	899
11ae80818_7952_c1_49						409	900
11ap20714_4960432_c3_97						410	901
11ap20714_5271967_c1_60						411	902
11ap20714_7227202_f3_43.aa						412	903
11ap20714_7227202_f3_43.aa						413	904
14ce31519_15635927_f3_15						414	905
14ee41924_23527267_c3_107						415	906
14ee41924_23834800_f2_32						416	907
hp1p13939_25397327_f3_22						417	908
hp2e10911_24855312_c1_69						418	909
hp2e10911_3349_c1_63						419	910
hp4e13394_35957200_f1_21						420	911
hp4e13394_5964452_c2_97						421	912
hp4e53394_22864682_c2_86.aa						422	913
hp5e15044_4554652_f3_3						423	914
hp5p15575_29300311_c1_29						424	915
hp5p15575_33445317_f2_20.aa						425	916
hp5p15575_6140713_f2_18						426	917
hp5p15641_12195281_c1_24						427	918
hp5p15641_24304527_c3_35						428	919
hp5p15641_25635452_c3_34						429	920
hp5p15870_14350428_f1_1						430	921
hp6p10590_23440913_c2_31						431	922
hp6p10606_19546933_c3_31						432	923
hp6p10903_4398263_f3_6.aa						433	924
hp6p10903_4398263_f3_6.aa						434	925
hp6p10904_2214676_c1_14						435	926
hp6p10904_23704412_f2_5						436	927
01cp20708_36134808_f2_11						437	928
02ae31010_12504512_f3_28.aa						438	929
02ae31010_16833312_f2_19						439	930

- 86 -

02ae31010_2117087_f3_34						440	931
02ae31010_30208317_f1_14						441	932
02ae31010_36132785_f2_29						442	933
02ae31010_5085162_c1_47						443	934
02cp10615_26573462_c1_45						444	935
03ae10804_12609533_c1_26						445	936
03ae10804_21698400_c2_32						446	937
04ep41903_26757937_f3_16						447	938
04ep41903_26757937_f3_16						448	939
04ep41903_4101593_f2_10.aa						449	940
05ep10815_26570332_c2_99						450	941
05ep10815_4195292_c1_84						451	942
05ep10815_4719175_c1_83						452	943
06cp30603_679218_f2_34						453	944
06ep10615_961562_f2_41						454	945
06ep11202_133293_c1_19						455	946
06ep11202_26353438_c1_22						456	947
06ep11202_4884677_c1_17						457	948
06ep11202_792962_c2_26.aa						458	949
06ep11917_24803153_c3_24						459	950
06ep30223_16512_c3_160						460	951
06ep30223_23476067_c1_119						461	952
06gp71906_15115637_f2_59						462	953
06gp71906_25478192_c1_131						463	954
06gp71906_25504187_f3_112						464	955
09cp61003_492187_c2_80.aa						465	956
11ae80818_19632781_c3_57						466	957
11ae80818_7290627_c2_51						467	958
11ap20714_34023312_f3_46						468	959
11ee11408_4977193_c1_41.aa						469	960
11ge10308_5256_f2_1						470	961
12ap10324_13178562_f3_6						471	962
14ce61516_13073577_f2_12						472	963
14ee41924_16282067_c1_72						473	964
hp2e10911_4882027_c2_87						474	965
hp3e11188_47327_f2_5						475	966
hp3e11188_5082842_f3_12						476	967
hp4e13394_3368767_c1_80						477	968
hp5p15212_6928132_c3_34						478	969
hp5p15641_30273312_c2_28						479	970
hp5p15641_5211687_c2_29						480	971
hp6p10590_30521093_f2_14						481	972
hp6p10904_7089062_c1_16						482	973
hp6p12129_16603417_f3_14						483	974
hp6p12244_3948467_c1_52						484	975
hp6p22217_23470967_f1_4						485	976
hp7e10192_4412568_f2_5						486	977
hp7p10287_24611325_c2_24						487	978
hp7p10290_25548812_f3_14						488	979
hp7p10290_25585941_f3_12						489	980
hp7p10290_35156558_f3_15						490	981
hp7p10290_4351718_f1_6						491	982

- 87 -

11ae11922_12586675_f2_1						983	1037
02ge10116_23866562_c3_146						984	1038
07ee11402_2458267_c3_108						985	1039
07ee50709_10213593_f3_77						986	1040
06ep10615_14649077_f2_30						987	1041
09cp10713_23452_c3_195						988	1042
11ap20714_7227202_f3_40						989	1043
01ce61016_23609580_c3_139						990	1044
06gp71906_3024126_c1_128						991	1045
06ge20501_14100018_c1_34						992	1046
09cp10713_34024187_f1_31						993	1047
hp4p62853_5914693_c3_52						994	1048
13ae10610_859692_c2_32						995	1049
13ae10610_35912_f2_3						996	1050
06ap11119_24426508_f3_27						997	1051
14gp11423_26803801_f3_7						998	1052
06ap10609_12586675_f2_19						999	1053
07ee50709_4818967_f2_43						1000	1054
02ap11117_23495187_c3_81						1001	1055
09cp10713_34024187_f1_31						1002	1056
07ee50709_35156577_f3_80						1003	1057
05ae30220_976413_c3_204						1004	1058
hp3e11188_47327_f2_9						1005	1059
12ap10324_4805318_f2_6						1006	1060
hpl80523_23485968_c2_49						1007	1061
hp7e10192_25598277_c2_15						1008	1062
hp4e53394_26209843_c3_98						1009	1063
06ep10615_9842_f1_5						1010	1064
06ep30223_34409437_f2_64						1011	1065
06gp10409_3398427_f2_12						1012	1066
09ce10413_5865665_f1_4						1013	1067
09cp10224_1062966_c1_44						1014	1068
01ce61016_12931513_c2_106						1015	1069
01ce61016_23609580_c3_139						1016	1070
14cp11908_25593768_c3_97						1017	1071
14cp11908_783127_c1_72						1018	1072
07ee11402_10759567_c2_86						1019	1073
07ee50709_4818967_f2_43						1020	1074
hp4e13394_5088562_f3_54						1021	1075
hp4e13394_15828963_c2_90						1022	1076
hp4e53394_19720300_c3_98						1023	1077
07ce10312_4554652_f3_2						1024	1078
05ae30220_14350428_f3_91						1025	1079
hp8e10080_19546933_c2_88						1026	1080
01ce10320_30273587_f3_38						1027	1081
07ee50709_26438968_f2_36						1028	1082
05ep10815_4719175_c1_115						1029	1083
06ep30223_23476067_c1_115						1030	1084
hp7e10590_26172564_c1_68						1031	1085
05ae30220_4977193_c3_198						1032	1086
hp7e10557_21698387_f1_1						1033	1087
07ee11402_19565702_c2_88						1034	1088

- 88 -

07ee50709_960952_f2_47						1035	1089
hp6p12244_3948467_c3_88						1036	1090
hp7e10590_13073577_c3_107						1293	1296
01ce61016_492187_c3_120						1294	1297
06ep10615_961562_f1_15						1295	1298

### EXEMPLIFICATION

#### 5 I. Cloning and Sequencing of *H. pylori* DNA

*H. pylori* chromosomal DNA was isolated according to a basic DNA protocol outlined in Schleif R.F. and Wensink P.C., *Practical Methods in Molecular Biology*, p.98, Springer-Verlag, NY., 1981, with minor modifications. Briefly, cells were pelleted, resuspended in TE (10 mM Tris, 1 mM EDTA, pH 7.6) and GES lysis buffer (5.1 M guanidium thiocyanate, 0.1 M EDTA, pH 8.0, 0.5% N-laurylsarcosine) was added. Suspension was chilled and ammonium acetate (NH<sub>4</sub>Ac) was added to final concentration of 2.0 M. DNA was extracted, first with chloroform, then with phenol-chloroform, and reextracted with chloroform. DNA was precipitated with isopropanol, washed twice with 70% EtOH, dried and resuspended in TE.

Following isolation whole genomic *H. pylori* DNA was nebulized (Bodenteich et al., *Automated DNA Sequencing and Analysis* (J.C. Venter, ed.), Academic Press, 1994) to a median size of 2000 bp. After nebulization, the DNA was concentrated and separated on a standard 1% agarose gel. Several fractions, corresponding to approximate sizes 900-1300 bp, 1300-1700 bp, 1700-2200 bp, 2200-2700 bp, were excised from the gel and purified by the GeneClean procedure (Bio101, Inc.).

The purified DNA fragments were then blunt-ended using T4 DNA polymerase. The healed DNA was then ligated to unique BstXI-linker adopters in 100-1000 fold molar excess. These linkers are complimentary to the BstXI-cut pMPX vectors, while the overhang is not self-complimentary. Therefore, the linkers will not concatemerize nor will the cut-vector religate itself easily. The linker-adopted inserts were separated from the unincorporated linkers on a 1% agarose gel and purified using GeneClean. The linker-adopted inserts were then ligated to each of the 20 pMPX vectors to construct a series of "shotgun" subclone libraries. The vectors contain an out-of-frame lacZ gene at the cloning site which becomes in-frame in the event that an adapter-dimer is cloned, allowing these to be avoided by their blue-color.

All subsequent steps were based on the multiplex DNA sequencing protocols outlined in Church G.M. and Kieffer-Higgins S., *Science* 240:185-188, 1988. Only major modifications to the protocols are highlighted. Briefly, each of the 20 vectors was

- 89 -

then transformed into DH5 $\alpha$  competent cells (Gibco/BRL, DH5 $\alpha$  transformation protocol). The libraries were assessed by plating onto antibiotic plates containing ampicillin, methicillin and IPTG/Xgal. The plates were incubated overnight at 37°C. Successful transformants were then used for plating of clones and pooling into the  
5 multiplex pools. The clones were picked and pooled into 40 ml growth medium cultures. The cultures were grown overnight at 37°C. DNA was purified using the Qiagen Midi-prep kits and Tip-100 columns (Qiagen, Inc.). In this manner, 100  $\mu$ g of DNA was obtained per pool. 15 96-well plates of DNA were generated to obtain a 5-10 fold sequence redundancy with 250-300 base average read-lengths.

10 These purified DNA samples were then sequenced using the multiplex DNA sequencing based on chemical degradation methods (Church G.M. and Kieffer-Higgins S., *Science* 240:185-188, 1988) or by Sequithrem (Epicenter Technologies) dideoxy sequencing protocols. The sequencing reactions were electrophoresed and transferred onto nylon membranes by direct transfer electrophoresis from 40 cm gels (Richterich P.  
15 and Church G.M., *Methods in Enzymology* 218:187-222, 1993) or by electroblotting (Church, *supra*). 24 samples were run per gel. 45 successful membranes were produced by chemical sequencing and 8 were produced by dideoxy sequencing. The DNA was covalently bound to the membranes by exposure to ultraviolet light, and hybridized with labeled oligonucleotides complimentary to tag sequences on the vectors (Church, *supra*).  
20 The membranes were washed to rinse off non-specifically bound probe, and exposed to X-ray film to visualize individual sequence ladders. After autoradiography, the hybridized probe was removed by incubation at 65° C, and the hybridization cycle repeated with another tag sequence until the membrane has been probed 38 times for chemical sequencing membranes and 10 times for the dideoxy sequencing membranes.  
25 Thus, each gel produced a large number of films, each containing new sequencing information. Whenever a new blot was processed, it was initially probed for an internal standard sequence added to each of the pools.

Digital images of the films were generated using a laser-scanning densitometer (Molecular Dynamics, Sunnyvale, CA). The digitized images were processed on  
30 computer workstations (VaxStation 4000's) using the program REPLICA™ (Church et al., *Automated DNA Sequencing and Analysis* (J.C. Venter, ed.), Academic Press, 1994). Image processing included lane straightening, contrast adjustment to smooth out intensity differences, and resolution enhancement by iterative gaussian deconvolution. The sequences were then automatically picked in REPLICA™ and displayed for  
35 interactive proofreading before being stored in a project database. The proofreading was accomplished by a quick visual scan of the film image followed by mouse clicks on the

- 90 -

bands of the displayed image to modify the base calls. For typical sequences derived by chemical sequencing, the error rate of the REPLICA™ base calling software was 2-5% with most errors occurring near the end of a sequence read. Many of the sequence errors could be detected and corrected because multiple sequence reads covering the same portion of the genomic DNA provide adequate sequence redundancy for editing. Each sequence automatically received a number correspond to (microtiter plate and probe information) and lane set number (corresponding to microtiter plate columns). This number serves as a permanent identifier of the sequence so it is always possible to identify the original of any particular sequence without recourse to a specialized database.

Routine assembly of *H. pylori* sequences was done using the program FALCON (Church, Church et al., *Automated DNA Sequencing and Analysis* (J.C. Venter, ed.), Academic Press, 1994). This program has proven to be fast and reliable for most sequences. The assembled contigs were displayed using a modified version of GelAssemble, developed by the Genetics Computer Group (GCG) (Devereux et al., *Nucleic Acid Res.* 12:387-95, 1984) that interacts with REPLICA™. This provided for an integrated editor that allows multiple sequence gel images to be instantaneously called up from the REPLICA™ database and displayed to allow rapid scanning of contigs and proofreading of gel traces where discrepancies occurred between different sequence reads in the assembly.

## II. Identification, Cloning and Expression of *H. pylori* Nucleic Acids

Expression and purification of the *H. pylori* polypeptides of the invention can be performed essentially as outlined below.

To facilitate the cloning, expression and purification of membrane and secreted proteins from *H. pylori*, a gene expression system, such as the pET System (Novagen), for cloning and expression of recombinant proteins in *E. coli*, is selected. Also, a DNA sequence encoding a peptide tag, the His-Tag, is fused to the 3' end of DNA sequences of interest in order to facilitate purification of the recombinant protein products. The 3' end is selected for fusion in order to avoid alteration of any 5' terminal signal sequence. The exception to the above is *ppiB*, a gene cloned for use as a control in the expression studies. The sequence for *H. pylori ppiB* contains a DNA sequence encoding a His-Tag fused to the 5' end of the full length gene, because the protein product of this gene does not contain a signal sequence and is expressed as a cytosolic protein.

- 91 -

*PCR Amplification and Cloning of Nucleic Acids Containing ORF's for Membrane and Secreted Polypeptides from H. pylori*

Nucleic acids chosen (for example, from the nucleic acids set forth in the Sequence Listing) for cloning from the J99 strain of *H. pylori* are prepared for  
5 amplification cloning by polymerase chain reaction (PCR). Synthetic oligonucleotide primers specific for the 5' and 3' ends of open reading frames (ORFs) are designed and purchased from GibcoBRL Life Technologies (Gaithersburg, MD, USA). All forward primers (specific for the 5' end of the sequence) are designed to include an NcoI cloning site at the extreme 5' terminus. These primers are designed to permit initiation of  
10 protein translation at a methionine residue followed by a valine residue and the coding sequence for the remainder of the native *H. pylori* DNA sequence. All reverse primers (specific for the 3' end of any *H. pylori* ORF) include a EcoRI site at the extreme 5' terminus to permit cloning of each *H. pylori* sequence into the reading frame of the pET-28b. The pET-28b vector provides sequence encoding an additional 20 carboxy-  
15 terminal amino acids including six histidine residues (at the extreme C-terminus), which comprise the His-Tag. An exception to the above, as noted earlier, is the vector construction for the *ppiB* gene. A synthetic oligonucleotide primer specific for the 5' end of *ppiB* gene encodes a BamHI site at its extreme 5' terminus and the primer for the 3' end of the *ppiB* gene encodes a XhoI site at its extreme 5' terminus.

20 Genomic DNA prepared from the J99 strain of *H. pylori* is used as the source of template DNA for PCR amplification reactions (Current Protocols in Molecular Biology, John Wiley and Sons, Inc., F. Ausubel et al., eds., 1994). To amplify a DNA sequence containing an *H. pylori* ORF, genomic DNA (50 nanograms) is introduced into a reaction vial containing 2 mM MgCl<sub>2</sub>, 1 micromolar synthetic oligonucleotide primers  
25 (forward and reverse primers) complementary to and flanking a defined *H. pylori* ORF, 0.2 mM of each deoxynucleotide triphosphate; dATP, dGTP, dCTP, dTTP and 2.5 units of heat stable DNA polymerase (Amplitaq, Roche Molecular Systems, Inc., Branchburg, NJ, USA) in a final volume of 100 microliters.

Upon completion of thermal cycling reactions, each sample of amplified DNA is  
30 washed and purified using the Qiaquick Spin PCR purification kit (Qiagen, Gaithersburg, MD, USA). All amplified DNA samples are subjected to digestion with the restriction endonucleases, e.g., NcoI and EcoRI (New England BioLabs, Beverly, MA, USA)(Current Protocols in Molecular Biology, John Wiley and Sons, Inc., F. Ausubel et al., eds., 1994). DNA samples are then subjected to electrophoresis on 1.0 %  
35 NuSeive (FMC BioProducts, Rockland, ME USA) agarose gels. DNA is visualized by exposure to ethidium bromide and long wave uv irradiation. DNA contained in slices

- 92 -

isolated from the agarose gel is purified using the Bio 101 GeneClean Kit protocol (Bio 101 Vista, CA, USA)

*Cloning of H. pylori Nucleic Acids Into an Expression Vector*

5 The pET-28b vector is prepared for cloning by digestion with endonucleases, e.g., NcoI and EcoRI (Current Protocols in Molecular Biology, John Wiley and Sons, Inc., F. Ausubel et al., eds., 1994). In the case of cloning ppiB, the pET-28a vector, which encodes a His-Tag that can be fused to the 5' end of an inserted gene, is used and the cloning site prepared for cloning with the ppiB gene by digestion with BamHI and XhoI restriction endonucleases.

10 Following digestion, DNA inserts are cloned (Current Protocols in Molecular Biology, John Wiley and Sons, Inc., F. Ausubel et al., eds., 1994) into the previously digested pET-28b expression vector, except for the amplified insert for ppiB, which is cloned into the pET-28a expression vector. Products of the ligation reaction are then used to transform the BL21 strain of *E. coli* (Current Protocols in Molecular Biology, 15 John Wiley and Sons, Inc., F. Ausubel et al., eds., 1994) as described below.

*Transformation Of Competent Bacteria With Recombinant Plasmids*

Competent bacteria, *E. coli* strain BL21 or *E. coli* strain BL21(DE3), are transformed with recombinant pET expression plasmids carrying the cloned *H. pylori* sequences according to standard methods (Current Protocols in Molecular, John Wiley and Sons, Inc., F. Ausubel et al., eds., 1994). Briefly, 1 microliter of ligation reaction is mixed with 50 microliters of electrocompetent cells and subjected to a high voltage pulse, after which, samples are incubated in 0.45 milliliters SOC medium (0.5% yeast extract, 2.0 % tryptone, 10 mM NaCl, 2.5 mM KCl, 10 mM MgCl<sub>2</sub>, 10 mM MgSO<sub>4</sub> and 20, mM glucose) at 37°C with shaking for 1 hour. Samples are then spread on LB agar 25 plates containing 25 microgram/ml kanamycin sulfate for growth overnight. Transformed colonies of BL21 are then picked and analyzed to evaluate cloned inserts as described below.

*Identification Of Recombinant Expression Vectors With H. Pylori Nucleic Acids*

30 Individual BL21 clones transformed with recombinant pET-28b-H.pylori ORFs are analyzed by PCR amplification of the cloned inserts using the same forward and reverse primers, specific for each *H. pylori* sequence, that were used in the original PCR amplification cloning reactions. Successful amplification verifies the integration of the *H. pylori* sequences in the expression vector (Current Protocols in Molecular Biology, John Wiley and Sons, Inc., F. Ausubel et al., eds., 1994).

- 93 -

### *Isolation and Preparation of Nucleic Acids From Transformants*

Individual clones of recombinant pET-28b vectors carrying properly cloned *H. pylori* ORFs are picked and incubated in 5 mls of LB broth plus 25 microgram/ml kanamycin sulfate overnight. The following day plasmid DNA is isolated and purified  
5 using the Qiagen plasmid purification protocol (Qiagen Inc., Chatsworth, CA, USA).

### *Expression Of Recombinant H. Pylori Sequences In E. coli*

The pET vector can be propagated in any *E. coli* K-12 strain e.g. HMS174, HB101, JM109, DH5, etc. for the purpose of cloning or plasmid preparation. Hosts for expression include *E. coli* strains containing a chromosomal copy of the gene for T7  
10 RNA polymerase. These hosts are lysogens of bacteriophage DE3, a lambda derivative that carries the lacI gene, the lacUV5 promoter and the gene for T7 RNA polymerase. T7 RNA polymerase is induced by addition of isopropyl-B-D-thiogalactoside (IPTG), and the T7 RNA polymerase transcribes any target plasmid, such as pET-28b, carrying its gene of interest. Strains used include: BL21(DE3) (Studier, F.W., Rosenberg, A.H.,  
15 Dunn, J.J., and Dubendorff, J.W. (1990) Meth. Enzymol. 185, 60-89).

To express recombinant *H. pylori* sequences, 50 nanograms of plasmid DNA isolated as described above is used to transform competent BL21(DE3) bacteria as described above (provided by Novagen as part of the pET expression system kit). The lacZ gene (beta-galactosidase) is expressed in the pET-System as described for the *H.*  
20 *pylori* recombinant constructions. Transformed cells are cultured in SOC medium for 1 hour, and the culture is then plated on LB plates containing 25 micrograms/ml kanamycin sulfate. The following day, bacterial colonies are pooled and grown in LB medium containing kanamycin sulfate (25 micrograms/ml) to an optical density at 600 nm of 0.5 to 1.0 O.D. units, at which point, 1 millimolar IPTG was added to the culture  
25 for 3 hours to induce gene expression of the *H. pylori* recombinant DNA constructions.

After induction of gene expression with IPTG, bacteria are pelleted by centrifugation in a Sorvall RC-3B centrifuge at 3500 x g for 15 minutes at 4°C. Pellets are resuspended in 50 milliliters of cold 10 mM Tris-HCl, pH 8.0, 0.1 M NaCl and 0.1 mM EDTA (STE buffer). Cells are then centrifuged at 2000 x g for 20 min at 4°C. Wet  
30 pellets are weighed and frozen at -80°C until ready for protein purification.

### III. Purification Of Recombinant Proteins From E. Coli

#### *Analytical Methods*

The concentrations of purified protein preparations are quantified  
35 spectrophotometrically using absorbance coefficients calculated from amino acid content (Perkins, S.J. 1986 Eur. J. Biochem. 157, 169-180). Protein concentrations are

also measured by the method of Bradford, M.M. (1976) Anal. Biochem. 72, 248-254, and Lowry, O.H., Rosebrough, N., Farr, A.L. & Randall, R.J. (1951) J. Biol. Chem. 193, pages 265-275, using bovine serum albumin as a standard.

SDS-polyacrylamide gels (12% or 4.0 to 25 % acrylamide gradient gels) are purchased from BioRad (Hercules, CA, USA), and stained with Coomassie blue. Molecular weight markers include rabbit skeletal muscle myosin (200 kDa), *E. coli* (-galactosidase (116 kDa), rabbit muscle phosphorylase B (97.4 kDa), bovine serum albumin (66.2 kDa), ovalbumin (45 kDa), bovine carbonic anhydrase (31 kDa), soybean trypsin inhibitor (21.5 kDa), egg white lysozyme (14.4 kDa) and bovine aprotinin (6.5 kDa).

#### 1. Purification of soluble proteins

All steps are carried out at 4°C. Frozen cells are thawed, resuspended in 5 volumes of lysis buffer (20 mM Tris, pH 7.9, 0.5 M NaCl, 5 mM imidazole with 10% glycerol, 0.1 % -mercaptoethanol, 200 (g/ ml lysozyme, 1 mM phenylmethylsulfonyl fluoride (PMSF), and 10 ug/ml each of leupeptin, aprotinin, pepstatin, L-1-chloro-3-[4-tosylamido]-7-amino-2-heptanone (TLCK), L-1-chloro-3-[4-tosylamido]-4-phenyl-2-butanone (TPCK), and soybean trypsin inhibitor, and ruptured by several passages through a small volume microfluidizer (Model M-110S, Microfluidics International Corporation, Newton, MA). The resultant homogenate is made 0.1 % Brij 35, and centrifuged at 100,000 x g for 1 hour to yield a clear supernatant (crude extract).

Following filtration through a 0.8 (m Supor filter (Gelman Sciences, FRG) the crude extract is loaded directly onto a Ni<sup>2+</sup>- nitrolotriacetate-agarose (NTA) with a 5 milliliter bed volume (Hochuli, E., Dbeli, H., and Schacheer, A. (1987) J. Chromatography 411, 177-184) pre-equilibrated in lysis buffer containing 10 % glycerol, 0.1 % Brij 35 and 1 mM PMSF. The column is washed with 250 ml (50 bed volumes) of lysis buffer containing 10 % glycerol, 0.1 % Brij 35, and are eluted with sequential steps of lysis buffer containing 10 % glycerol, 0.05 % Brij 35, 1 mM PMSF, and: either 20, 100, 200, or 500 mM imidazole. Fractions are monitored by absorbance at OD<sub>280</sub> nm, and peak fractions are analyzed by SDS-PAGE.

#### 2. Purification of insoluble proteins from inclusion bodies

The following steps are carried out at 4°C. Cell pellets are resuspended in lysis buffer with 10% glycerol 200 (g/ ml lysozyme, 5 mM EDTA, 1mM PMSF and 0.1 % -mercaptoethanol. After passage through the cell disrupter, the resulting homogenate is made 0.2 % deoxycholate, stirred 10 minutes, then centrifuged at 20,000 x g, for 30 min. The pellets are washed with lysis buffer containing 10 % glycerol, 10 mM EDTA, 1% Triton X-100, 1 mM PMSF and 0.1% -mercaptoethanol, followed by several washes

- 95 -

with lysis buffer containing 1 M urea, 1 mM PMSF and 0.1 % -mercaptoethanol. The resulting white pellet is composed primarily of inclusion bodies, free of unbroken cells and membranous materials.

*Dialysis and concentration of protein samples*

5 Urea is removed slowly from the protein samples by dialysis against Tris-buffered saline (TBS; 10 mM Tris pH 8.0, 150 mM NaCl) containing 0.5 % deoxycholate (DOC) with sequential reduction in urea as follows; 6M, 4M, 3M, 2M, 1M, 0.5 M and finally TBS without any urea. Each dialysis step is conducted for a minimum of 4 hours at room temperature.

10 After dialysis, samples are concentrated by pressure filtration using Amicon stirred-cells. Protein concentrations are measured using the methods of Perkins (1986 Eur. J. Biochem. 157, 169-180), Bradford ((1976) Anal. Biochem. 72, 248-254) and Lowry ((1951) J. Biol. Chem. 193, pages 265-275).

15 IV. Assessment Of The Antigenicity Of Outer Membrane Localized Antigens Of *H. pylori*

Purification of outer membranes form *H. pylori* can be performed by essentially following the protocol outlined below.

20 *H. pylori* strains J99 (ATCC# 55679) and Ah244 are grown on chocolate blood agar containing 5% (vol/vol) horse blood, at 37°C in an atmosphere containing 10% CO<sub>2</sub> for 48 h. Bacteria were harvested by suspension in 20 mM Tris, pH 7.5. The cells are collected by centrifugation at 12,000 Xg, for 20 min at 4°C and washed 3 times with 20 mM Tris, pH 7.5. Cells are suspended in 20 mM Tris, pH 7.5 and broken by sonication on ice (eight bursts of 30 s at 60 watts with 60 s pauses between bursts).

25 DNase (0.1 mg) and RNase (0.5 mg) are added to the cell suspension, and the mixture is incubated for 30 minutes at room temperature. The cell suspension is centrifuged at 12,000 Xg for 20 min, at 4°C. The supernatant was retained and centrifuged again. Total membranes are collected from the supernatant by centrifugation at 40,000 Xg for 30 minute, at 4°C. The pellet are washed twice in 20 mM Tris, pH 7.5. The protein

30 content is assayed using the Bradford protein assay, with bovine serum albumin (BSA) as a standard. The suspension is then adjusted to 1 mg protein /ml. The solubilization of the membranes is realized by adding N-lauryl-sarcosine to this suspension in a ratio of 6 mg of N-lauryl-sarcosine per mg of protein. The suspension is incubated for 30 minutes at room temperature in presence of N-lauryl-sarcosine. Outer membranes are

35 collected by centrifugation at 40,000 Xg for 30 minutes at 4°C. The pellet is washed 3 times with Milli Q quality water, aliquoted and stored at -20°C until use.

- 96 -

*Identification Of Outer Membrane Antigens of H. pylori*

Outer membrane antigens can be identified using a protocol outlined below.

Proteins are separated on sodium dodecylsulfate polyacrylamide gels (SDS-PAGE) according to the method described by Laemmli, U.K. (1970) Nature (London) Volume 227, 680-685. Samples are prepared by suspension in standard treatment buffer and heated at 100°C for 10 min. Approximately 1-5 mg of protein is loaded per well on 8X10 cm minigels (0.75 mm). The separated proteins are then transferred to PVDF membranes as described below.

Electroblotting of separated proteins to PVDF membranes is performed in a Bio Rad Mini-Trans Blot Electrophoretic Transfer cell. The PVDF membrane Immobilon-PSQ is employed. Electroblotting is carried out for 60 min at 50V using CAPS transfer buffer (10mM 3-[Cyclohexylamino]-1-propanesulfonic acid, 10% methanol). The membrane is stained with 0.2% Ponceau S and destained with Milli Q quality water.

Antigens within the preparation are then identified using western immunoblotting. After electroblotting, non specific binding sites of the PVDF membrane are blocked with 5% non fat dry milk in 10 mM Tris-HCl-0.9% NaCl, pH 7.5. The membrane is incubated with a appropriate dilution of normal mouse serum in 10 mM Tris-HCl-0.9% NaCl-0.5% Tween 20-0.5% BSA, pH 7.5, for 2 h at 37°C and then washed three times with 10 mM Tris-HCl-0.9% NaCl-0.5% Tween 20, pH 7.5 (TTBS). Alkaline phosphatase conjugated anti-mouse Ig, from goat is then added in 10 mM Tris-HCl-0.9% NaCl-0.5% Tween 20-0.5% BSA, pH 7.5 and incubated for 1h at room temperature. After this incubation, the membrane is washed three times in TTBS. The reactive bands are revealed using 5-bromo-4-chloro-3-indolyl phosphate (Bio-Rad) as the Alkaline phosphatase substrate and Nitro Blue Tetrazolium (Bio-Rad) as the color development reagent.

For amino acid microsequencing, proteins that are identified as immunoreactive are cut from a fresh unreacted immobilon membranes and microsequenced at the Worcester Foundation microsequencing facility. Membranes from which the protein bands are cut are then subjected to western immunoblot as described above to confirm that the appropriate band had been excised.

V. Analysis Of *H. Pylori* Proteins As Vaccine Candidates

To investigate the immunomodulatory effect of *H. pylori* proteins, a mouse/*H. pylori* model was used. This model mimics the human *H. pylori* infection in many respects. The focus is on the effect of oral immunization in *H. pylori* infected animals in order to test the concept of therapeutic oral immunotherapy.

- 97 -

*Animals*

Female SPF BALB/c mice were purchased from Bomholt Breeding center (Denmark). They were kept in ordinary makrolon cages with free supply of water and food. The animals were 4-6 weeks old at arrival.

5      *Infection*

After a minimum of one week of acclimatization, the animals were infected with a type 2 strain (VacA negative) of *H. pylori* (strain 244, originally isolated from an ulcer patient). In our hands, this strain has earlier proven to be a good colonizer of the mouse stomach. The bacteria were grown overnight in Brucella broth supplemented with 10 % fetal calf serum, at 37°C in a microaerophilic atmosphere (10% CO<sub>2</sub>, 5% O<sub>2</sub>). The animals were given an oral dose of omeprazole (400 µmol/kg) and 3-5 h after this an oral inoculation of *H. pylori* in broth (approximately 10<sup>8</sup> cfu/animal). Positive take of the infection was checked in some animals 2-3 weeks after the inoculation.

*Antigens*

15      Recombinant *H. pylori* antigens were chosen based on their association with externally exposed *H. pylori* cell membrane. These antigens were selected from the following groups: (1.) Outer Membrane Proteins; (2.) Periplastic/Secreted proteins; (3.) Outer Surface proteins; and (4.) Inner Membrane proteins. All recombinant proteins were constructed with a hexa-HIS tag for purification reasons and the non-*Helicobacter*  
20      *pylori* control protein (β-galactosidase from *E. coli*; LacZ), was constructed in the same way.

All antigens were given in a soluble form, i.e. dissolved in either a HEPES buffer or in a buffer containing 0.5% Deoxycholate (DOC).

25      The antigens are listed in Table 8 below.

Table 8*Helicobacter pylori* proteins**Outer membrane Proteins**

30      Protein 1  
Protein 2  
Protein 3  
Protein 4  
Protein 5

35      **Periplastic/Secreted proteins**  
Protein 6

40      **Other cell envelope proteins**  
Protein 7  
Protein 8

- 98 -

### Flagella-associated proteins Protein 9

### Control proteins

5  $\beta$ -galactosidase (LacZ)

### Immunizations

Ten animals in each group were immunized 4 times over a 34 day period (day 1, 15, 25 and 35). Purified antigens in solution or suspension were given at a dose of 100  $\mu$ g/mouse. As an adjuvant, the animals were also given 10  $\mu$ g/mouse of Cholera toxin (CT) with each immunization. Omeprazole (400  $\mu$ mol/kg) was given orally to the animals 3-5 h prior to immunization as a way of protecting the antigens from acid degradation. Infected control animals received HEPES buffer + CT or DOC buffer + CT. Animals were sacrificed 2-4 weeks after final immunization. A general outline of the study is shown in Table 9 below.

Table 9

### Study outline, therapeutic immunization:

20 Mice were all infected with *H. pylori* strain Ah244 at day 30.

	<u>Substance</u>	<u>Mouse strain</u> <u>n=10</u>	<u>Dose/mouse</u>	<u>Dates for</u> <u>dosing</u>
25	1. Controls, PBS	Balb/c	0.3 ml	0, 14, 24, 34
	2. Cholera toxin, 10 $\mu$ g	Balb/c	0.3 ml	0, 14, 24, 34
	3. Protein 1, 100 $\mu$ g + CT 10 $\mu$ g	Balb/c	0.3 ml	0, 14, 24, 34
	4. Protein 5, 100 $\mu$ g + CT 10 $\mu$ g	Balb/c	0.3 ml	0, 14, 24, 34
	5. Protein 10, 100 $\mu$ g + CT 10 $\mu$ g	Balb/c	0.3 ml	0, 14, 24, 34
30	6. Protein 9, 100 $\mu$ g + CT 10 $\mu$ g	Balb/c	0.3 ml	0, 14, 24, 34
	7. Protein 2, 100 $\mu$ g + CT 10 $\mu$ g	Balb/c	0.3 ml	0, 14, 24, 34
	8. Protein 6, 100 $\mu$ g + CT 10 $\mu$ g	Balb/c	0.3 ml	0, 14, 24, 34
	9. Protein 4, 100 $\mu$ g + CT 10 $\mu$ g	Balb/c	0.3 ml	0, 14, 24, 34
	10. Protein 7, 100 $\mu$ g + CT 10 $\mu$ g	Balb/c	0.3 ml	0, 14, 24, 34
35	11. Protein 8, 100 $\mu$ g + CT 10 $\mu$ g	Balb/c	0.3 ml	0, 14, 24, 34
	12. Protein 3, 100 $\mu$ g + CT 10 $\mu$ g	Balb/c	0.3 ml	0, 14, 24, 34

### Analysis of infection

Mucosal infection: The mice were sacrificed by CO<sub>2</sub> and cervical dislocation. The abdomen was opened and the stomach removed. After cutting the stomach along the greater curvature, it was rinsed in saline. The mucosa from the antrum and corpus of an area of 25mm<sup>2</sup> was scraped separately with a surgical scalpel. The mucosa scraping was suspended in Brucella broth and plated onto Blood Skirrow selective plates. The

- 99 -

plates were incubated under microaerophilic conditions for 3-5 days and the number of colonies was counted. The identity of *H. pylori* was ascertained by urease and catalase test and by direct microscopy or Gram staining.

5 The urease test was performed essentially as follows. The reagent, Urea Agar Base Concentrate, was purchased from DIFCO Laboratories, Detroit, MI (Catalog # 0284-61-3). Urea agar base concentrate was diluted 1:10 with water. 1 ml of if the diluted concentrate was mixed with 100-200  $\mu$ l of actively growing *H. pylori* cells. Color change to magenta indicated that cells were urease positive.

10 The catalase test was performed essentially as follows. The reagent, N,N,N',N'-Tetramethyl-p-Phenylenediamine, was purchased from Sigma, St. Louis, MO (Catalog # T3134). A solution of the reagent (1% w/v in water) was prepared. *H. pylori* cells were swabbed onto Whatman filter paper and overlaid with the 1% solution. Color change to dark blue indicated that the cells were catalase positive.

15 Serum antibodies: From all mice serum was prepared from blood drawn by heart puncture. Serum antibodies were identified by regular ELISA techniques, where the specific antigens of *Helicobacter pylori* were plated.

20 Mucosal antibodies: Gentle scrapings of a defined part of the corpus and of 4 cm of duodenum were performed in 50% of the mice in order to detect the presence of antibodies in the mucous. The antibody titers were determined by regular ELISA technique as for serum antibodies.

Statistical analysis: Wilcoxon-Mann-Whitney sign rank test was used for determination of significant effects of the antigens on *Helicobacter pylori* colonization.  $P < 0.05$  was considered significant. Because the antrum is the major colonization site for *Helicobacter* most emphasis was put upon changes in the antral colonization.

25

### Results

Antibodies in sera: All antigens tested given together with CT gave rise to a measurable specific titer in serum. The highest responses were seen with Proteins 3, 4, 9, 1, and 7 (see Figure 1).

30 Antibodies in mucus: In the mucus scrapings, specific antibodies against all antigens tested were seen. By far the strongest response was seen with Protein 6, followed by 1, 3, and 9 (see Figure 2).

### Therapeutic immunization effects:

35 All control animals (BALB/c mice) were well colonized with *H. pylori* (strain AH244) in both antrum and corpus of the stomach. Of the antigens tested 3 proteins (Proteins 4, 7, and 1) gave a good and significant reduction and/or eradication of the *H.*

- 100 -

*pylori* infection. The degree of colonization of the antrum was lower following immunization with Proteins 8, 9, and 3 compared to control. The effect of Proteins 5, 2, and 6 did not differ from control. The control protein lacZ, i.e. the non-*H. pylori* protein, had no eradication effect and in fact had higher *Helicobacter* colonization compared to the HEPES + CT control. All data are shown in Figures 3 and 4 for proteins dissolved in HEPES and DOC respectively. Data is shown as geometric mean values. n=8-10 Wilcoxon-Mann-Whitney sign rank test \* = p<0.05; x/10 = number of mice showing eradication of *H. pylori* over the total number of mice examined.

The data presented indicate that all of the *H. pylori* associated proteins included in this study, when used as oral immunogens in conjunction with the oral adjuvant CT, resulted in stimulation of an immune response as measured by specific serum and mucosal antibodies. A majority of the proteins led to a reduction, and in some cases complete clearance of the colonization of *H. pylori* in this animal model. It should be noted that the reduction or clearance was due to heterologous protection rather than homologous protection (the polypeptides were based on the *H. pylori* J99 strain sequence and used in the therapeutic immunization studies against a different (AH244) challenge strain), indicating the vaccine potential against a wide variety of *H. pylori* strains.

The highest colonization in the antrum was seen in animals treated with the non-*Helicobacter* protein LacZ, indicating that the effects seen with the *Helicobacter pylori* antigens were specific.

Taken together these data strongly support the use of these *H. pylori* proteins in a pharmaceutical formulation for the use in humans to treat and/or prevent *H. pylori* infections.

#### VI. Sequence Variance Analysis of genes in *Helicobacter pylori* strains

Four genes were cloned and sequenced from several strains of *H. pylori* to compare the DNA and deduced amino acid sequences. This information was used to determine the sequence variation between the *H. pylori* strain, J99, and other *H. pylori* strains isolated from human patients.

##### *Preparation of Chromosomal DNA.*

Cultures of *H. pylori* strains (as listed in Table 12) were grown in BLBB (1% Tryptone, 1% Peptamin 0.1% Glucose, 0.2% Yeast Extract 0.5% Sodium Chloride, 5% Fetal Bovine Serum) to an OD<sub>600</sub> of 0.2. Cells were centrifuged in a Sorvall RC-3B at 3500 x g at 4°C for 15 minutes and the pellet resuspended in 0.95 mls of 10 mM Tris-HCl, 0.1 mM EDTA (TE). Lysozyme was added to a final concentration of 1mg/ml

- 101 -

along with, SDS to 1% and RNase A + T1 to 0.5mg/ml and 5 units/ml respectively, and incubated at 37°C for one hour. Proteinase K was then added to a final concentration of 0.4mg/ml and the sample was incubated at 55 C for more than one hour. NaCl was added to the sample to a concentration of 0.65 M, mixed carefully, and 0.15 ml of 10% CTAB in 0.7M NaCL (final is 1% CTAB/70mM NaCL) was added followed by incubation at 65°C for 20 minutes. At this point, the samples were extracted with chloroform:isoamyl alcohol, extracted with phenol, and extracted again with chloroform:isoamyl alcohol. DNA was precipitated with either EtOH (1.5 x volumes) or isopropanol (0.6 x volumes) at -70°C for 10minutes, washed in 70% EtOH and resuspended in TE.

#### *PCR Amplification and cloning.*

Genomic DNA prepared from twelve strains of *Helicobacter pylori* was used as the source of template DNA for PCR amplification reactions (Current Protocols in Molecular Biology, John Wiley and Sons, Inc., F. Ausubel et al., editors, 1994). To amplify a DNA sequence containing an *H. pylori* ORF, genomic DNA (10 nanograms) was introduced into a reaction vial containing 2 mM MgCl<sub>2</sub>, 1 micromolar synthetic oligonucleotide primers (forward and reverse primers, see Table 10) complementary to and flanking a defined *H. pylori* ORF, 0.2 mM of each deoxynucleotide triphosphate; dATP, dGTP, dCTP, dTTP and 0.5 units of heat stable DNA polymerase (Amplitaq, Roche Molecular Systems, Inc., Branchburg, NJ, USA) in a final volume of 20 microliters in duplicate reactions.

Table 10

Oligonucleotide primers used for PCR amplification of *H. pylori* DNA sequences.

Outer membrane Proteins	Forward primer 5' to 3'	Reverse Primer 5' to 3'
Protein 11 (for strains AH4, AH15, AH61, 5294, 5640, AH18, and AH244)	5'-TTAACCATGGTGAAAAGC GATA-3' (SEQ ID NO:1091)	5'-TAGAATTCGCCTCTAAAA CTTAG-3' (SEQ ID NO:1092)
Protein 11 (for strains AH5, 5155, 7958, AH24, and J99)	5'-TTAACCATGGTGAAAAGC GATA-3' (SEQ ID NO:1093)	5'-TAGAATTCGCATAACGAT CAATC-3' (SEQ ID NO:1094)
Protein 12	5'-ATATCCATGGTGAGTTTG ATGA-3' (SEQ ID NO:1095)	5'-ATGAATTCAATTTTTTATT TTGCCA-3' (SEQ ID NO:1096)
Protein 13	5'-AATTCCATGGCTATCCAA ATCCG-3' (SEQ ID NO:1097)	5'-ATGAATTCGCCAAAATC GTAGTATT-3' (SEQ ID NO:1098)
Protein 14	5'-GATACCATGGAATTTATGA AAAAG-3' (SEQ ID NO:1099)	5'-TGAATTCGAAAAAGTGT AGTTATAC-3' (SEQ ID NO:1100)

- 102 -

The following thermal cycling conditions were used to obtain amplified DNA products for each ORF using a Perkin Elmer Cetus/ GeneAmp PCR System 9600 thermal cycler:

- 5 Sequences for Proteins 12 and 14;  
Denaturation at 94°C for 2 min,  
2 cycles at 94°C for 15 sec, 30°C for 15 sec and 72°C for 1.5 min  
23 cycles at 94°C for 15 sec, 55°C for 15 sec and 72°C for 1.5 min  
Reactions were concluded at 72°C for 6 minutes.
- 10 Sequence for Protein 11 for strains AH5, 5155, 7958, AH24, and J99;  
Denaturation at 94°C for 2 min,  
2 cycles at 94°C for 15 sec, 30°C for 15 sec and 72°C for 1.5 min  
25 cycles at 94°C for 15 sec, 55°C for 15 sec and 72°C for 1.5 min  
15 Reaction was concluded at 72°C for 6 minutes.

- Sequences for Protein 11 and Protein 13 for strains AH4, AH15, AH61, 5294, 5640, AH18, and Hp244 ;  
Denaturation at 94°C for 2 min,  
20 2 cycles at 94°C for 15 sec, 30°C for 20 sec and 72°C for 2 min  
25 cycles at 94°C for 15 sec, 55°C for 20 sec and 72°C for 2 min  
Reactions were concluded at 72°C for 8 minutes.

- Upon completion of thermal cycling reactions, each pair of samples were  
25 combined and used directly for cloning into the pCR cloning vector as described below.  
*Cloning of H. pylori DNA sequences into the pCR TA cloning vector.*

- All amplified inserts were cloned into the pCR 2.1 vector by the method described in the Original TA cloning kit (Invitrogen, San Diego, CA). Products of the ligation reaction were then used to transform the TOP10F' (INVaF' in the case of *H. pylori* sequence 350) strain of *E. coli* as described below.
- 30

*Transformation of competent bacteria with recombinant plasmids*

- Competent bacteria, *E. coli* strain TOP10F' or *E. coli* strain INVaF' were transformed with recombinant pCR expression plasmids carrying the cloned *H. pylori* sequences according to standard methods (Current Protocols in Molecular Biology, John Wiley and Sons, Inc., F. Ausubel et al., editors, 1994). Briefly, 2 microliters of 0.5 micromolar BME was added to each vial of 50 microliters of competent cells. Subsequently, 2 microliters of ligation reaction was mixed with the competent cells and incubated on ice for 30 minutes. The cells and ligation mixture were then subjected to a "heat shock" at 42°C for 30 seconds, and were subsequently placed on ice for an additional 2 minutes, after which, samples were incubated in 0.45 milliliters SOC medium (0.5% yeast extract, 2.0 % tryptone, 10 mM NaCl, 2.5 mM KCl, 10 mM MgCl<sub>2</sub>, 10 mM MgSO<sub>4</sub> and 20, mM glucose) at 37°C with shaking for 1 hour. Samples
- 35  
40

- 103 -

were then spread on LB agar plates containing 25 microgram/ml kanamycin sulfate or 100 micrograms/ml ampicillin for growth overnight. Transformed colonies of TOP10F' or INVaF' were then picked and analyzed to evaluate cloned inserts as described below.

*Identification of recombinant PCR plasmids carrying H. pylori sequences*

- 5 Individual TOP10F' or INVaF' clones transformed with recombinant pCR-*H. pylori* ORFs were analyzed by PCR amplification of the cloned inserts using the same forward and reverse primers, specific for each *H. pylori* sequence, that were used in the original PCR amplification cloning reactions. Successful amplification verified the integration of the *H. pylori* sequences in the cloning vector (Current Protocols in
- 10 Molecular Biology, John Wiley and Sons, Inc., F. Ausubel et al., editors, 1994).

- Individual clones of recombinant pCR vectors carrying properly cloned *H. pylori* ORFs were picked for sequence analysis. Sequence analysis was performed on ABI Sequencers using standard protocols (Perkin Elmer) using vector-specific primers (as found in PCR11 or pCR2.1, Invitrogen, San Diego, CA) and sequencing primers specific
- 15 to the ORF as listed in Table 11 below.

**Table 11**

Oligonucleotide primers used for sequencing of *H. pylori* DNA sequences.

Outer membrane Proteins	Forward primers 5' to 3'	Reverse Primers 5' to 3'
<b>Protein 11</b>	5'-CCCTTCATTTTAGAAATCG-3' (SEQ ID NO:1101) 5'-ATTTCAACCAATTCAATGCG-3' (SEQ ID NO:1102) 5'-GCCCCTTTTGATTGAAGCT-3' (SEQ ID NO:1103) 5'-TCGCTCCAAGATACCAAG AAGT-3' (SEQ ID NO:1104) 5'-CTTGAATTAGGGGCAAAG ATCG-3' (SEQ ID NO:1105) 5'-ATGCGTTTTTACCCAAAGA AGT-3' (SEQ ID NO:1106) 5'-ATAACGCCACTTCCTTATTG GT-3' (SEQ ID NO:1107)	5'-CTTTGGGTAAAAACGCATC-3' (SEQ ID NO:1108) 5'-CGATCTTTGATCCTAATTCA-3' (SEQ ID NO:1109) 5'-ATCAAGTTGCCTATGCTGA-3' (SEQ ID NO:1110)
<b>Protein 12</b>	5'-TTGAACACTTTTGATTATG CGG-3' (SEQ ID NO:1111) 5'-GGATTATGCGATTGTTTCA CAAG-3' (SEQ ID NO:1112)	5'-GTCTTTAGCAAAAATGGCGTC- 3' (SEQ ID NO:1113) 5'-AATGAGCGTAAGAGAGCCTTC- 3' (SEQ ID NO:1114)
<b>Protein 13</b>	5'-CTTATGGGGGTATTGTCA-3' (SEQ ID NO:1115) 5'-AGCATGTGGGTATCCAGC-3' (SEQ ID NO:1116)	5'-AGGTTGTTGCCTAAAGACT-3' (SEQ ID NO:1117) 5'-CTGCCTCCACCTTTGATC-3' (SEQ ID NO:1118)

- 104 -

<b>Protein 14</b>	5'-ACCAATATCAATTGGCACT-3' (SEQ ID NO:1119) 5'-ACTTGAAAAGCTCTGCA-3' (SEQ ID NO:1120)	5'-CTTGCTTGTCATATCTAGC-3' (SEQ ID NO:1121) 5'-GTTGAAGTGTTGGTGCTA-3' (SEQ ID NO:1122)
	5'-CAAGCAAGTGGTTTGGTTTT AG-3' (SEQ ID NO:1123) 5'-TGGAAAGAGCAAATCATTG AAG-3' (SEQ ID NO:1124)	5'-GCCCCATAATCAAAAAGCCCAT- 3' (SEQ ID NO:1125) 5'-CTAAAACCAAACCACTTGC TTGTC-3' (SEQ ID NO:1126)
<b>Vector Primers</b>	5'-GTAAAACGACGGCCAG-3' (SEQ ID NO:1127)	5'-CAGGAAACAGCTATGAC-3' (SEQ ID NO:1128)

### Results

To establish the PCR error rate in these experiments, five individual clones of Protein 11, prepared from five separate PCR reaction mixtures from *H. pylori* strain J99, were sequenced over a total length of 897 nucleotides for a cumulative total of 4485 bases of DNA sequence. DNA sequence for the five clones was compared to the DNA sequence of Protein 11 obtained previously by a different method, i.e., random shotgun cloning and sequencing. The PCR error rate for the experiments described herein was determined to be 2 base changes out of 4485 bases, which is equivalent to an estimated error rate of less than or equal to 0.04%.

DNA sequence analysis was performed on four different open reading frames identified as genes and amplified by PCR methods from a dozen different strains of the bacterium *Helicobacter pylori*. The deduced amino acid sequences of three of the four open reading frames that were selected for this study showed statistically significant BLAST homology to defined proteins present in other bacterial species. Those ORFs included: Protein 11, homologous to the val A & B genes encoding an ABC transporter in *F. novicida*; Protein 12, homologous to lipoprotein e (P4) present in the outer membrane of *H. influenzae*; Protein 13, homologous to fecA, an outer membrane receptor in iron (III) dicitrate transport in *E. coli*. Protein 14 was identified as an unknown open reading frame, because it showed low homology with sequences in the public databases.

To assess the extent of conservation or variance in the ORFs across various strains of *H. pylori*, changes in DNA sequence and the deduced protein sequence were compared to the DNA and deduced protein sequences found in the J99 strain of *H. pylori* (see Table 12 below). Results are presented as percent identity to the J99 strain of *H. pylori* sequenced by random shotgun cloning. To control for any variations in the J99 sequence each of the four open reading frames were cloned and sequenced again from the J99 bacterial strain and that sequence information was compared to the sequence information that had been collected from inserts cloned by random shotgun sequencing of the J99 strain. The data demonstrate that there is variation in the DNA

- 105 -

sequence ranging from as little as 0.12 % difference (Protein 14, J99 strain) to approximately 7% change (Protein 11, strain AH5). The deduced protein sequences show either no variation (Protein 14, strains AH18 and AH24) or up to as much as 7.66% amino acid changes (Protein 11, Strain AH5).

5

TABLE 12

Multiple Strain DNA Sequence analysis of *H. pylori* Vaccine Candidates

<u>J99 Protein #:</u>	<u>11</u>	<u>11</u>	<u>12</u>	<u>12</u>	<u>13</u>	<u>13</u>	<u>14</u>	<u>14</u>
Length of Region Sequenced:	248 a.a.	746 nt.	232 a.a.	696 nt.	182 a.a.	548 nt.	273 a.a.	819 nt.
<u>Strain Tested</u>	AA	Nuc.	AA	Nuc.	AA	Nuc.	AA	Nuc.
	identity	identity	identity	identity	identity	identity	identity	identity
J99	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	99.63%	99.88%
AH244	95.16%	95.04%	n.d.	n.d.	99.09%	96.71%	98.90%	96.45%
AH4	95.97%	95.98%	97.84%	95.83%	n.d.	n.d.	97.80%	95.73%
AH5	92.34%	93.03%	98.28%	96.12%	98.91%	96.90%	98.53%	95.73%
AH15	95.16%	94.91%	97.41%	95.98%	99.82%	97.99%	99.63%	96.09%
AH61	n.d.	n.d.	97.84%	95.98%	99.27%	97.44%	n.d.	n.d.
5155	n.d.	n.d.	n.d.	n.d.	99.45%	97.08%	98.53%	95.60%
5294	94.35%	94.37%	98.28%	95.40%	99.64%	97.26%	97.07%	95.48%
7958	94.35%	94.10%	97.84%	95.40%	n.d.	n.d.	99.63%	96.46%
5640	95.16%	94.37%	97.41%	95.69%	99.09%	97.63%	98.53%	95.48%
AH18	n.d.	n.d.	98.71%	95.69%	99.64%	97.44%	100.00%	95.97%
AH24	94.75%	95.04%	97.84%	95.40%	99.27%	96.71%	100.00%	96.46%

10

n.d. = not done

#### VII. Experimental Knock-Out Protocol for the Determination of Essential *H. pylori* Genes as Potential Therapeutic Targets

15

Therapeutic targets were chosen from genes whose protein products appear to play key roles in essential cell pathways such as cell envelope synthesis, DNA synthesis, transcription, translation, regulation and colonization/virulence.

20

The protocol for the deletion of portions of *H. pylori* genes/ORFs and the insertional mutagenesis of a kanamycin-resistance cassette was modified from previously published methods (Labigne-Roussel et al., 1988, J. Bacteriology 170, pp. 1704-1708; Cover et al., 1994, J. Biological Chemistry 269, pp. 10566-10573; Reytrat et al., 1995, Proc. Natl. Acad. Sci. 92, pp 8768-8772).

#### *Identification and Cloning of *H. pylori* Gene Sequences*

25

The sequences of the genes or ORFs (open reading frames) selected as knock-out targets were identified from the *H. pylori* genomic sequence and used to design primers to specifically amplify the genes/ORFs. All synthetic oligonucleotide primers (Table

- 106 -

13) were designed with the aid of the OLIGO program (National Biosciences, Inc., Plymouth, MN 55447, USA), and were purchased from Gibco/BRL Life Technologies (Gaithersburg, MD, USA). Specific primers (F1 and R1) were chosen which flanked most or all of the ORF, depending on its size. If the ORF was smaller than 800 to 1000  
5 base pairs, flanking primers were chosen outside of the open reading frame.

**TABLE 13**  
**Oligonucleotide Sequences for Knock-Out Gene/ORFs**

Gene Name	Accession Number	Cloning Primers	Deletion-creating Primers	Targeting Primers	
	F1	R1	F2	R2	R3
nhh	P23329	TTGCCCATCGTA TTGATAGA (SEQ ID NO:1129)	AGAGCGTATTCA CCCGAAAG (SEQ ID NO:1130)	TCTTGCATCTTAAT CCACTCC (SEQ ID NO:1131)	CGGGTCAAAACGA CCACTTAA (SEQ ID NO:1132)
ppib	P29820	TGGTATAAGGATT TGAATGGA (SEQ ID NO:1135)	TTGACTAAACACA TGCAGAA (SEQ ID NO:1136)	ATAGAGAGCGTTG TGTTTAGC (SEQ ID NO:1137)	CCTTTATTGGTTTT GATCGTG (SEQ ID NO:1138)
tsf	P34828	GCCTTTGGCTTCTT CGTTGTC (SEQ ID NO:1141)	GAAATGGAAAATA GCGGTCAA (SEQ ID NO:1142)	GCAAATCCCCAGC CACTTCC (SEQ ID NO:1143)	GTGGCTAAAAATG AGGGCTT (SEQ ID NO:1144)
MurD	P14900	GTTGGGCAGAAAA TAAGGTGA (SEQ ID NO:1147)	CAACAAACCTGA CAAGAAAC (SEQ ID NO:1148)	CATTGATGCCTAA AACTTCG (SEQ ID NO:1149)	CGTGGTGGTTTTTC CCGTTAG (SEQ ID NO:1150)
MurE	P22188	GCGTTTGGGGATT TGATGTTT (SEQ ID NO:1153)	GCGCGTAGAGGCT TGTAATA (SEQ ID NO:1154)	GCCCTGATCCATT CCCCCTT (SEQ ID NO:1155)	CTGTTTTTAGCGTC CCTGTA (SEQ ID NO:1156)
AlgA	P07874	GCGTTTGTGATTCT GTCTGTTA (SEQ ID NO:1159)	GTAAAAACACCCGC TAACGCAT (SEQ ID NO:1160)	GCGTGTCTTCTAA GGGTTC (SEQ ID NO:1161)	GGAATTTAAGCG TCTTTTT (SEQ ID NO:1162)
metL	P19358	GCCCCAGCCCCAT AATACAAA (SEQ ID NO:1165)	GGAGGGGGCAATT AAACATCG (SEQ ID NO:1166)	TCACGCTTCTTAA ATCATCA (SEQ ID NO:1167)	CGCTAATCACATC CTTTCTT (SEQ ID NO:1168)

TABLE 13 (continued)

flaA	X16278	GAATGCGGTGGT TTAGAGAG (SEQ ID NO:1171)	CGGTTTTTAAGAC TGAATACA (SEQ ID NO:1172)	GGCGATGTGATT GGCGATT (SEQ ID NO:1173)	GGATAGCCTGCCA AAACGCC (SEQ ID NO:1174)	AAGTTATGCGGG CGAGATT (SEQ ID NO:1175)	GGAGCAATCAGC CATTTTC (SEQ ID NO:1176)
FlgE	U09549	CTAGCGATTCAAG GCGATGG (SEQ ID NO:1177)	CGGCCTCCTTCAA ACACATT (SEQ ID NO:1178)	AGCGGGCAGTTTA GGACCAC (SEQ ID NO:1179)	GCATTGATCGCAT TTTTAGCC (SEQ ID NO:1180)	ACGGTTAGCAGG GCAGAA (SEQ ID NO:1181)	CAAAAGAGGCGG GTTCAATGC (SEQ ID NO:1182)
Flim- set A	M37691	TTTAGAAGTCGTT GATGAGA (SEQ ID NO:1183)	CATACACGCTCAC TTCATCG (SEQ ID NO:1184)	AGTGTGTCGCCT GTGGTGGAG (SEQ ID NO:1185)	CCCCTAATAGTCT GTCAATCAT (SEQ ID NO:1186)	CAAAAGATTGAAG CAGAAGAGT (SEQ ID NO:1187)	AATGGTTTTCCTA TACCCTTGA (SEQ ID NO:1188)
Flim- set B	M37691	GAGAGCAAAATCCT TATCCAG (SEQ ID NO:1189)	CATACACGCTCAC TTCATCG (SEQ ID NO:1190)	AGTGTGTCGCCT GTGGTGGAG (SEQ ID NO:1191)	CCCCTAATAGTCT GTCAATCAT (SEQ ID NO:1192)	CAAAAGATTGAAG CAGAAGAGT (SEQ ID NO:1193)	AATGGTTTTCCTA TACCCTTGA (SEQ ID NO:1194)
MurC	U32794	TTGAAACCCCAAA AGTTTAC (SEQ ID NO:1195)	TCAACTGATAGGT AATATCCC (SEQ ID NO:1196)	GCTAGGATTTATG CCAATTA (SEQ ID NO:1197)	TACGAGACAAAT AGGGATT (SEQ ID NO:1198)	ATAGGCATGCAGA ATTTTCC (SEQ ID NO:1199)	CCATTACATTCG CCTC (SEQ ID NO:1200)
dnaE	M10040	CGATAGATATTGT AGAAGTCA (SEQ ID NO:1201)	GGGCTTGATTCA TTTTGTAA (SEQ ID NO:1202)	GTTTAAACCAACGC CATAGCCA (SEQ ID NO:1203)	TTCTAAAGGTGG TAATCTTC (SEQ ID NO:1204)	TAAGTCAAGCCAT AAACCCAA (SEQ ID NO:1205)	TTTTGGGTAAAA AGGCTGAA (SEQ ID NO:1206)
SerS	X05017	ATCTTTTGCCCTT GCTCATA (SEQ ID NO:1207)	AGACAGCACCCAGT TTGATAAA (SEQ ID NO:1208)	CAGCCACACITCA ATGCTAT (SEQ ID NO:1209)	GTAAGGCGTTAGA AAATACC (SEQ ID NO:1210)	GCCCCATTAAAT CCTTTCT (SEQ ID NO:1211)	AAAGGATACAAG GGGA (SEQ ID NO:1212)
gly	P00960	CTCGCTCCATTTTA TCTTTA (SEQ ID NO:1213)	TTTTTAGGGAGG ATTGAGAT (SEQ ID NO:1214)	TGTTTGAAATGC TGGTGATC (SEQ ID NO:1215)	CTTTGGGGGAGT TTGACAA (SEQ ID NO:1216)	TTTGATAAACGCC CACTTTT (SEQ ID NO:1217)	TTTCAAAACGCTC ACCTTTG (SEQ ID NO:1218)
Glx	L14580	TCTATCTTTTGAT GCTCTCT (SEQ ID NO:1219)	ATAATGAGTTTGA TCGTTACG (SEQ ID NO:1220)	ACAATAATAGGCT TTGCTTC (SEQ ID NO:1221)	AATTAGCCCTTA AATAGAT (SEQ ID NO:1222)	AACAACCGCTAAA ATCAAC (SEQ ID NO:1223)	CTTCAGCGATACT AAAAGAT (SEQ ID NO:1224)
Sig28 (flhA)	M37691	TAGGGCGATTGA AACAGC (SEQ ID NO:1225)	GCTGGATAAGGAT TTGCTCT (SEQ ID NO:1226)	TTTTGGGGGTAT GCTAAA (SEQ ID NO:1227)	GGCTGGTAAATAC TGGATAG (SEQ ID NO:1228)	AGGCTATTCAGG TGGCTAAA (SEQ ID NO:1229)	ATTCTCATCAACG ACTTCTAAA (SEQ ID NO:1230)

TABLE 13 (continued)

Sig54	M73443	GCAGTTGGCGGTA TTTGGTG (SEQ ID NO:1231)	GAGAGCGAAGTTT ATGAGAA (SEQ ID NO:1232)	TGATTGTTGGGTA GCTCTCA (SEQ ID NO:1233)	AAATCGGTCTGA TGCTCTTA (SEQ ID NO:1234)	CTTTCTCTTCGCT TGAAGA (SEQ ID NO:1235)	AAAACAACGCA TCAAAAAAT (SEQ ID NO:1236)
MurI	U12405	TTTCAAGGCGAGG AGGCAGAT (SEQ ID NO:1237)	GCACAAAGACCCC ACCACGAT (SEQ ID NO:1238)	CGCCCGAATGGAT GAGTAGG (SEQ ID NO:1239)	TGCAACAAAAATA CGCCCTT (SEQ ID NO:1240)	TTTTTAAGGCGGT ATTTTGT (SEQ ID NO:1241)	TGGGTTTTAAGGA ATGTGATG (SEQ ID NO:1242)
dnaB	D26185	CGCGCTCAAAATC CCTAAAT (SEQ ID NO:1243)	GGCCATTCTTTC GGATATT (SEQ ID NO:1244)	GCCCATTCCTGTT TTTAGC (SEQ ID NO:1245)	CGCTTAAAGCTC CTTTCAC (SEQ ID NO:1246)	AGCGTTTTTGTA GGGGGTAT (SEQ ID NO:1247)	TCCTATCATAGC GTTAGTGC (SEQ ID NO:1248)
MurG	D10602	CTTAGGGGTTTTT AGCATGAA (SEQ ID NO:1249)	GCACAAATCCAC ACGCTGC (SEQ ID NO:1250)	CCAAAGCTAAAGC GGTGT (SEQ ID NO:1251)	GCTCATGGATATA AAGGGTATT (SEQ ID NO:1252)	CTTAGCCCTTTA GTGTTTA (SEQ ID NO:1253)	CGCAAAAGGGTA GGGGATAA (SEQ ID NO:1254)
IpxC	U32794	TTTTATTTTAGAA ACGAATC (SEQ ID NO:1255)	CAAACTATCGCC CTCTCTA (SEQ ID NO:1256)	AAAGATAACGCTA GGATTCTAC (SEQ ID NO:1257)	TAATCTACAGAG TGTTAATGG (SEQ ID NO:1258)	GCGTCATGGAAT TTTTAGA (SEQ ID NO:1259)	ATTCAAAGAAAG CTGGCTGTCT (SEQ ID NO:1260)
kdtA	M86305	GCTTGTGGGGGTT GTTTTAT (SEQ ID NO:1261)	GAACCCCTAAAA TGACAAAT (SEQ ID NO:1262)	ACCATGCTCATT ACGCTAGG (SEQ ID NO:1263)	GTAAGTTTGAGCG GCTAATTC (SEQ ID NO:1264)	AAAAAGAAAGAA GAACTCGTG (SEQ ID NO:1265)	AAAGATACTCCC CTGTGATTA (SEQ ID NO:1266)
IpxB	U09549	CAAAGAAACGCAA AATACAG (SEQ ID NO:1267)	GCATGGTATTCAG CGTTTTT (SEQ ID NO:1268)	GCAGCGGCACAGC GACTTTAG (SEQ ID NO:1269)	CGCCCCCAAAAG TCGCAGTA (SEQ ID NO:1270)	AAAGGTTGAAAC AAGAAATCT (SEQ ID NO:1271)	CACCTTGAGCGTTA GCAACAAT (SEQ ID NO:1272)
KO 24		GGGGTGTGAA TTGATAGA (SEQ ID NO:1273)	CGTAGGAGAAAG GAAGGAA (SEQ ID NO:1274)	CAAGGCGTTTTT TGGGGTAT (SEQ ID NO:1275)	GGGATTGTACAG GAAAAGAT (SEQ ID NO:1276)	GGAATACAATAAC GCA'AAAT (SEQ ID NO:1277)	GCCTTTTAGACA ACCTACT (SEQ ID NO:1278)
KO 26		CCCATAAATCAA ATCTCAAT (SEQ ID NO:1279)	GCTAGAAATGCCA TGAGAAAG (SEQ ID NO:1280)	GCGATTATGGGT ATTATTG (SEQ ID NO:1281)	TTATTGTGGAGTT GCTTGTA (SEQ ID NO:1282)	TATCGGCTCATC CTATTA (SEQ ID NO:1283)	CCCTAAATCCAA ATCAAGCAG (SEQ ID NO:1284)

- 110 -

Genomic DNA prepared from the *Helicobacter pylori* HpJ99 strain (ATCC 55679) was used as the source of template DNA for amplification of the ORFs by PCR (polymerase chain reaction) (Current Protocols in Molecular Biology, John Wiley and Sons, Inc., F. Ausubel et al., editors, 1994). For the preparation of genomic DNA from *H. pylori*, see

5 Example I. PCR amplification was carried out by introducing 10 nanograms of genomic HpJ99 DNA into a reaction vial containing 10 mM Tris pH 8.3, 50 mM KCl, 2 mM MgCl<sub>2</sub>, 2 microMolar synthetic oligonucleotide primers (forward=F1 and reverse=R1), 0.2 mM of each deoxynucleotide triphosphate (dATP, dGTP, dCTP, dTTP), and 1.25 units of heat stable DNA polymerase (Amplitaq, Roche Molecular Systems, Inc.,

10 Branchburg, NJ, USA) in a final volume of 40 microliters. The PCR was carried out with Perkin Elmer Cetus/GeneAmp PCR System 9600 thermal cyclers. The thermal cycling conditions used to obtain amplified DNA products for each knock-out target are shown in Table 14.

**TABLE 14**  
**PCR Conditions**

MurC

Denaturation at 94°C for 2 min.,  
25 cycles of 94°C for 15 sec, 48°C for 15 sec, 72°C for 1 min, 30 sec,  
20 Final Extension of 72°C for 20 minutes.

Sig54

Denaturation at 94°C for 2 min.,  
25 cycles of 94°C for 15 sec, 50°C for 15 sec, 72°C for 1 min, 30 sec,  
25 Final Extension of 72°C for 20 minutes.

lpxC

Denaturation at 94°C for 2 min.,  
32 cycles of 94°C for 15 sec, 50°C for 15 sec, 72°C for 1 min, 30 sec,  
30 Final Extension of 72°C for 20 minutes.

KO24, KO26, KO27

Denaturation at 94°C for 2 min.,  
25 cycles of 94°C for 15 sec, 50.5°C for 20 sec, 72°C for 2 min,  
35 Final Extension of 72°C for 20 minutes.

KO29, 26kDa Protein

Denaturation at 94°C for 2 min.,  
28 cycles of 94°C for 15 sec, 50.5°C for 20 sec, 72°C for 2 min,  
40 Final Extension of 72°C for 20 minutes.

DnaE, Glycyl

Denaturation at 94°C for 2 min.,  
20 cycles of 94°C for 15 sec, 51°C for 15 sec, 72°C for 1 min, 30 sec,  
45 Final Extension of 72°C for 20 minutes.

- 111 -

Gltx

Denaturation at 94°C for 2 min.,  
25 cycles of 94°C for 15 sec, 51°C for 15 sec, 72°C for 1 min, 30 sec,  
Final Extension of 72°C for 20 minutes.

KO28

Denaturation at 94°C for 2 min.,  
25 cycles of 94°C for 15 sec, 51°C for 15 sec, 72°C for 2min,  
Final Extension of 72°C for 20 minutes.

KO30

Denaturation at 94°C for 2 min.,  
25 cycles of 94°C for 15 sec, 51.5°C for 15 sec, 72°C for 1min, 45 sec,  
Final Extension of 72°C for 20 minutes.

MurI, MurG

Denaturation at 94°C for 2 min.,  
25 cycles of 94°C for 15 sec, 52°C for 15 sec, 72°C for 1 min, 30 sec,  
Final Extension of 72°C for 20 minutes.

DnaB, KdtA, LpxB

Denaturation at 94°C for 2 min.,  
27 cycles of 94°C for 15 sec, 52°C for 15 sec, 72°C for 1 min, 30 sec,  
Final Extension of 72°C for 20 minutes.

Tsf, FlgE, FliM, Sig28, MurB

Denaturation at 94°C for 2 min.,  
30 cycles of 94°C for 15 sec, 52°C for 15 sec, 72°C for 1 min, 30 sec,  
Final Extension of 72°C for 20 minutes.

PpiB

Denaturation at 94°C for 2 min.,  
30 cycles of 94°C for 15 sec, 52°C for 15 sec, 72°C for 2 min, 30 sec,  
Final Extension of 72°C for 20 minutes.

MurD, MurE, AlgA, MetL, FusA, SerS, Rnh

Denaturation at 94°C for 2 min.,  
30 cycles of 94°C for 15 sec, 55°C for 15 sec, 72°C for 1 min, 30 sec,  
Final Extension of 72°C for 20 minutes.

Upon completion of thermal cycling reactions, each sample of amplified DNA was visualized on a 2% TAE agarose gel stained with Ethidium Bromide (Current Protocols in Molecular Biology, John Wiley and Sons, Inc., F. Ausubel et al., editors, 1994) to determine that a single product of the expected size had resulted from the reaction. Amplified DNA was then washed and purified using the Qiaquick Spin PCR purification kit (Qiagen, Gaithersburg, MD, USA).

PCR products were cloned into the pT7Blue T-Vector (catalog#69820-1, Novagen, Inc., Madison, WI, USA) using the TA cloning strategy (Current Protocols in Molecular Biology, John Wiley and Sons, Inc., F. Ausubel et al., editors, 1994). The

- 112 -

ligation of the PCR product into the vector was accomplished by mixing a 6 fold molar excess of the PCR product, 10 ng of pT7Blue-T vector (Novagen), 1 microliter of T4 DNA Ligase Buffer (New England Biolabs, Beverly, MA, USA), and 200 units of T4 DNA Ligase (New England Biolabs) into a final reaction volume of 10 microliters.

5 Ligation was allowed to proceed for 16 hours at 16°C.

Ligation products were electroporated (Current Protocols in Molecular Biology, John Wiley and Sons, Inc., F. Ausubel et al., editors, 1994) into electroporation-competent XL-1 Blue or DH5- $\alpha$  *E. coli* cells (Clontech Lab., Inc. Palo Alto, CA, USA). Briefly, 1 microliter of ligation reaction was mixed with 40 microliters of

10 electrocompetent cells and subjected to a high voltage pulse (25 microFarads, 2.5 kV, 200 ohms) after which the samples were incubated in 0.45 ml SOC medium (0.5% yeast extract, 2% tryptone, 10 mM NaCl, 2.5 mM KCl, 10 mM MgCl<sub>2</sub>, 10 mM MgSO<sub>4</sub> and 20 mM glucose) at 37°C with shaking for 1 hour. Samples were then spread onto LB (10 g/l bacto tryptone, 5 g/l bacto yeast extract, 10 g/l sodium chloride) plates containing  
15 100 microgram/ml of Ampicillin, 0.3% X-gal, and 100 microgram/ml IPTG. These plates were incubated overnight at 37°C. Ampicillin-resistant colonies with white color were selected, grown in 5 ml of liquid LB containing 100 microgram/ml of Ampicillin, and plasmid DNA was isolated using the Qiagen miniprep protocol (Qiagen, Gaithersburg, MD, USA).

20 To verify that the correct *H. pylori* DNA inserts had been cloned, these pT7Blue plasmid DNAs were used as templates for PCR amplification of the cloned inserts, using the same forward and reverse primers (F1 and R1) used for the initial amplification of the J99 *H. pylori* sequence. Recognition of the primers and a PCR product of the correct size as visualized on a 2% TAE, ethidium bromide stained agarose gel were  
25 confirmation that the correct inserts had been cloned. Two to six such verified clones were obtained for each knock-out target, and frozen at -70°C for storage. To minimize errors due to PCR, plasmid DNA from these verified clones were pooled, and used in subsequent cloning steps.

The sequences of the genes/ORFs were again used to design a second pair of  
30 primers (F2 and R2) which flanked the region of *H. pylori* DNA to be either interrupted or deleted (up to 250 basepairs) within the ORFs but were oriented away from each other. The pool of circular plasmid DNAs of the previously isolated clones were used as templates for this round of PCR. Since the orientation of amplification of this pair of deletion primers was away from each other, the portion of the ORF between the primer  
35 would not be included in the resultant PCR product. The PCR product was a linear piece of DNA with *H. pylori* DNA at each end and the pT7Blue vector backbone

- 113 -

between them which, in essence, resulted in the deletion of a portion of the ORFs. The PCR product was visualized on a 1% TAE, ethidium bromide stained agarose gel to confirm that only a single product of the correct size had been amplified.

A Kanamycin-resistance cassette (Labigne-Roussel et al., 1988 J. Bacteriology 170, 1704-1708) was ligated to this PCR product by the TA cloning method used previously (Current Protocols in Molecular Biology, John Wiley and Sons, Inc., F. Ausubel et al., editors, 1994). The Kanamycin cassette containing a *Campylobacter* kanamycin resistance gene was obtained by carrying out an EcoRI digestion of the recombinant plasmid pCTB8:*kan* (Cover et al., 1994, J. Biological Chemistry 269, pp. 10566-10573). The proper fragment (1.4 kb) was isolated on a 1% TAE gel, and isolated using the QIAquick gel extraction kit (Qiagen, Gaithersburg, MD, USA). The fragment was end repaired using the Klenow fill-in protocol, which involved mixing 4 µg of the DNA fragment, 1 microliter of dATP, dGTP, dCTP, dTTP at 0.5 mM, 2 microliter of Klenow Buffer (New England Biolabs) and 5 units of Klenow DNA Polymerase I Large (Klenow) Fragment (New England Biolabs) into a 20 microliter reaction, incubating at 30°C for 15 min, and inactivating the enzyme by heating to 75°C for 10 minutes. This blunt-ended Kanamycin cassette was then purified through a Qiaquick column (Qiagen, Gaithersburg, MD, USA) to eliminate nucleotides. The 5' overhang was then generated by mixing 5 micrograms of the blunt-ended kanamycin cassette, 10 mM Tris pH 8.3, 50 mM KCl, 2 mM MgCl<sub>2</sub>, 5 units of DNA Polymerase (Amplitaq, Roche Molecular Systems, Inc., Branchburg, NJ, USA), 20 microliters of 5 mM dTTP, in a 100 microliter reaction and incubating the reaction for 2 hours at 37°C. The "Kan-T" cassette was purified using a QIAquick column (Qiagen, Gaithersburg, MD, USA). The PCR product of the deletion primers (F2 and R2) was ligated to the Kan-T cassette by mixing 10 to 25 ng of deletion primer PCR product, 50 - 75 ng Kan-T cassette DNA, 1 microliter 10x T4 DNA Ligase reaction mixture, 0.5 microliter T4 DNA Ligase (New England Biolabs, Beverly, MA, USA) in a 10 microliter reaction reaction and incubating for 16 hours at 16°C.

The ligation products were transformed into XL-1 Blue or DH5-α *E. coli* cells by electroporation as described previously. After recovery in SOC, cells were plated onto LB plates containing 100 microgram/ml Ampicillin and grown overnight at 37°C. These plates were then replica plated onto plates containing 25 microgram/ml Kanamycin and allowed to grow overnight. Resultant colonies had both the Ampicillin resistance gene present in the pT7Blue vector, and the newly introduced Kanamycin resistance gene. Colonies were picked into LB containing 25 microgram/ml Kanamycin and plasmid

- 114 -

DNA was isolated using the Qiagen miniprep protocol (Qiagen, Gaithersburg, MD, USA).

Several tests by PCR amplification were conducted on these plasmids to verify that the Kanamycin was inserted in the *H. pylori* gene/ORF, and to determine the orientation of the insertion of the Kanamycin-resistance gene relative to the *H. pylori* gene/ORF. To verify that the Kanamycin cassette was inserted into the *H. pylori* sequence, the plasmid DNAs were used as templates for PCR amplification with the set of primers (F1 and R1) originally used to clone the *H. pylori* gene/ORFs. The correct PCR product was the size of the deleted gene/ORF but increased in size by the addition of a 1.4 kilobase Kanamycin cassette. To avoid potential polar effects of the kanamycin resistance cassette on *H. pylori* gene expression, the orientation of the Kanamycin resistance gene with respect to the knock-out gene/ORF was determined and both orientations were eventually used in *H. pylori* transformations (see below). To determine the orientation of insertion of the kanamycin resistance gene, primers were designed from the ends of the kanamycin resistance gene ("Kan-1" 5'-ATCTTACCTATCACCTCAAAT-3' (SEQ ID NO:1285), and "Kan-2" 5'-AGACAGCAACATCTTTGTGAA-3' (SEQ ID NO:1286)). By using each of the cloning primers (F1 and R1) in conjunction with each of the Kan primers (4 combinations of primers), the orientation of the Kanamycin cassette relative to the *H. pylori* sequence was determined. Positive clones were classified as either in the "A" orientation (the same direction of transcription was present for both the *H. pylori* gene and the Kanamycin resistance gene), or in the "B" orientation (the direction of transcription for the *H. pylori* gene was opposite to that of the Kanamycin resistance gene). Clones which shared the same orientation (A or B) were pooled for subsequent experiments and independently transformed into *H. pylori*.

#### *Transformation of Plasmid DNA into H. pylori cells*

Two strains of *H. pylori* were used for transformation: ATCC 55679, the clinical isolate which provided the DNA from which *H. pylori* sequence database was obtained, and AH244, an isolate which had been passaged in, and had the ability to colonize the mouse stomach. Cells for transformation were grown at 37°C, 10% CO<sub>2</sub>, 100% humidity, either on Sheep-Blood agar plates or in Brucella Broth liquid. Cells were grown to exponential phase, and examined microscopically to determine that the cells were "healthy" (actively moving cells) and not contaminated. If grown on plates, cells were harvested by scraping cells from the plate with a sterile loop, suspended in 1 ml of Brucella Broth, spun down (1 minute, top speed in eppendorf microfuge) and

resuspended in 200 microliters Brucella Broth. If grown in Brucella Broth liquid, cells were centrifuged (15 minutes at 3000 rpm in a Beckman TJ6 centrifuge) and the cell pellet resuspended in 200 microliters of Brucella broth. An aliquot of cells was taken to determine the optical density at 600 nm, in order to calculate the concentration of cells.

- 5 An aliquot (1 to 5 OD<sub>600</sub> units/25 microliter) of the resuspended cells was placed onto a prewarmed Sheep-Blood agar plate, and the plate was further incubated at 37°C, 6% CO<sub>2</sub>, 100% humidity for 4 hours. After this incubation, 10 microliters of plasmid DNA (100 micrograms per microliter), was spotted onto these cells. A positive control (plasmid DNA with the ribonuclease H gene disrupted by kanamycin resistance gene)
- 10 and a negative control (no plasmid DNA) were done in parallel. The plates were returned to 37°C, 6% CO<sub>2</sub> for an additional 4 hours of incubation. Cells were then spread onto that plate using a swab wetted in Brucella broth, and grown for 20 hours at 37°C, 6% CO<sub>2</sub>. Cells were then transferred to a Sheep-Blood agar plate containing 25 micrograms/ml Kanamycin, and allowed to grow for 3 to 5 days at 37°C, 6% CO<sub>2</sub>,
- 15 100% humidity. If colonies appeared, they were picked and regrown as patches on a fresh Sheep-Blood agar plate containing 25 micrograms/ml Kanamycin.

- Three sets of PCR (three tests) were done to verify that the colonies of transformants have arose from homologous recombination at the proper chromosomal location. The template for PCR (DNA from the colony) was obtained by a rapid boiling
- 20 DNA preparation method. An aliquot of the colony (stab of the colony with a toothpick) was introduced into 100 microliters of 1% Triton X-100, 20 mM Tris, pH 8.5, and boiled for 6 minutes. An equal volume of phenol : chloroform (1:1) was added and vortexed. The mixture was microfuged for 5 minutes and the supernatant was used as DNA template for PCR with combinations of the following primers to verify
- 25 homologous recombination at the proper chromosomal location.

- TEST 1 PCR with F1 and R1 primers (cloning primers originally used to amplify the gene/ORF). A positive result of homologous recombination at the correct chromosomal location should show a single PCR product whose size is expected to be the size of the deleted gene/ORF but increased in size by the addition of a 1.4 kilobase
- 30 Kanamycin cassette. A PCR product of just the size of the gene/ORF was proof that the gene had not been knocked out and that the transformant was not the result of homologous recombination at the correct chromosome location.

- TEST 2 PCR with F3 (primer designed from sequences upstream of the gene/ORF), and either primer Kan-1 or Kan-2 (primers designed from the ends of the
- 35 kanamycin resistance gene), depending on whether the plasmid DNA used was of "A" or "B" orientation. A positive result of homologous recombination at the correct

- 116 -

chromosomal location of the sequences of the gene/ORFs upstream from the kanamycin resistance gene should show a single PCR product, the expected size to be from the location of F3 to the insertion site of kanamycin resistance gene. No PCR product or PCR product(s) of incorrect size(s) would prove that the plasmid had not been integrated at the correct site and that the gene had not been knocked out.

TEST 3 PCR with R3 (primer designed from sequences downstream of the gene/ORF) and either primer Kan-1 or Kan-2, depending on whether the plasmid DNA used was of "A" or "B" orientation. A positive result of homologous recombination at the correct chromosomal location downstream from the kanamycin resistance gene would show a single PCR product, the expected size to be from the insertion site of kanamycin resistance gene to the downstream location of R3. Again, no PCR product or PCR product(s) of incorrect size(s) would prove that the plasmid had not been integrated at the correct site and that the gene had not been knocked out.

Genes that are not essential for survival *in vitro* normally resulted in many transformants as observed for the positive control of ribonuclease H gene. Any transformants showing positive results for all three tests above would result in the conclusion that the gene was not essential for survival *in vitro*.

Genes that are essential for survival *in vitro* normally showed very few transformants. All transformants would be screened. A negative result of any of the three above tests for each transformant would lead to the conclusion that the gene had not been disrupted, and that the gene was essential for survival *in vitro*.

In the event that no colonies resulted from two independent transformations while the positive control with the disrupted ribonuclease H plasmid DNA produced transformants, the plasmid DNA was further analyzed by PCR on DNA from transformant populations prior to plating for colony formation, to verify that it can enter the cells and undergo homologous recombination at the correct site. Briefly, plasmid DNA was incubated according to the transformation protocol described above. DNA was extracted from the *H. pylori* cells immediately after incubation with the plasmid DNA and the DNA was used as templates for the above TEST 2 and TEST 3. Positive results of TEST 2 and TEST 3 would verify that the plasmid DNA could enter the cells and undergo homologous recombination at the correct chromosomal location. If TEST 2 and TEST 3 are positive, then failure to obtain viable transformants indicates that the gene is essential and cells suffering a disruption in that gene are incapable of colony formation

Genes used in these experiments have been found to be essential, non-essential, or are still in progress, as indicated in Table 15.

- 117 -

TABLE 15

Summary of knock-out genes/ORFs			
Gene	Accession Number	Pathway	Status
rnh	P23329	Transcription	Not essential
ppiB	P29820	Translation	Not essential
tsf	P34828	Translation	Essential
MurD	P14900	Cell envelope	Essential
MurE	P22188	Cell envelope	Essential
AlgA	P07874	Virulence/Colonization	Not essential
metL	P19358	Amino acid biosynthesis - aspartate family	Not essential
fusA	X16278	Translation	Essential
FlgE	U09549	Virulence/Colonization	Not essential-motility impaired
FliM	M37691	Virulence/Colonization	Not essential-motility impaired
MurC	U32794	Cell envelope	Essential
dnaE	M10040	DNA replication	Essential
serS	X05017	Translation	Essential
gly	P00960	Translation	Essential
gltX	L14580	Translation	Essential
sig28 (fliA)	M37691	Regulatory functions	Not essential-motility impaired
sig54		Regulatory functions	Not essential-motility impaired
MurI	U12405	Cell envelope	Essential
dnaB	D26185	DNA replication	In progress
MurG	D10602	Cell envelope	In progress
lpxC/envA	U32794	Cell envelope	In progress
kdtA	M86305	Cell envelope	In progress
lpxB	U09549	Cell envelope	In progress
KO 24		thiolase-like	In progress
KO 26		histone-like	In progress
KO 27		respiratory chain NADH dehydrogenase-like	In progress

VIII. Cloning, purification and characterization of the gene encoding the peptidyl-propyl *cis-trans* isomerase of *H. pylori*

The *Helicobacter pylori* genome contains an open reading frame (ORF) of 170 amino acids was found to have homology with the *Synechococcus sp.* (strain PCC 7942).ppi gene (NCBI Accession number P29820). Therefore, to evaluate whether this ORF encoded a protein with PPIase activity, the gene was isolated by polymerase chain reaction (PCR) amplification cloning, overexpressed in *E. coli*, and the protein purified to homogeneity. To facilitate purification, a polyhistidine tag was added to the N-terminus of this ORF. A simple assay using PPIase to evaluate protein folding function was developed for future use as a high-throughput drug screen.

Currently, the class of PPIases is divided into three unrelated families: the cyclophilins, the FK506-binding (FKBPs) and the parvulins. Although PPIase mutants have been reported from yeast and fruit fly, attempts to isolate disruption mutants in *Escherichia coli* were unsuccessful (Shieh, B.H., et.al. (1989) *Nature* 338:67-70). This suggests that this activity is essential for viability in bacteria.

*Cloning, expression and protein purification*

To facilitate the cloning, expression and purification of ppi from *H. pylori*, a powerful gene expression system, the pET System, for cloning and expression of recombinant ppi in *E. coli*. In this study, the sequence for *H. pylori* ppi contains a DNA sequence encoding a His-Tag fused to the 5' end of the full length gene, because the protein product of this gene does not contain a signal sequence and is expressed as a cytosolic protein.

A synthetic oligonucleotide primer (5'-TTATGGATCCAAACCAATTAAAA CT-3' (SEQ ID NO:1287)) specific for the 5' end of ppi gene encoded a BamHI site at its extreme 5' terminus and a primer (5'-TATCTCGAGTTATAGAGAAGGGC-3' (SEQ ID NO:1288)) specific for the 3' end of the ppi gene encoded a XhoI site at its extreme 5' terminus. Genomic DNA prepared from the J99 strain of *Helicobacter pylori* was used as the source of template DNA for PCR amplification reactions (Current Protocols in Molecular Biology, John Wiley and Sons, Inc., F. Ausubel et al., editors, 1994). To amplify a DNA sequence containing the *H. pylori* ppi gene, genomic DNA (50 nanograms) was introduced into a reaction vial containing 2 mM MgCl<sub>2</sub>, 1 micromolar synthetic oligonucleotide primers (forward and reverse primers) complementary to and flanking a defined *H. pylori* ORF, 0.2 mM of each deoxynucleotide triphosphate; dATP, dGTP, dCTP, dTTP and 2.5 units of heat stable DNA polymerase (Amplitaq, Roche Molecular Systems, Inc., Branchburg, NJ, USA) in a final volume of 100 microliters.

- 119 -

The following thermal cycling conditions were used to obtain amplified DNA products for each ORF using a Perkin Elmer Cetus/ GeneAmp PCR System 9600 thermal cycler:

*Conditions for amplification of H. pylori ppiB;*

Denaturation at 94°C for 2 min,  
2 cycles at 94°C for 15 sec, 32°C for 15 sec and 72°C for 1.5 min  
25 cycles at 94°C for 15 sec, 56°C for 15 sec and 72°C for 1.5 min  
Reactions were concluded at 72°C for 6 minutes.

Upon completion of thermal cycling reactions, the amplified DNA was washed and purified using the Qiaquick Spin PCR purification kit (Qiagen, Gaithersburg, MD, USA). The amplified DNA sample was subjected to digestion with the restriction endonucleases, BamHI and XhoII (New England BioLabs, Beverly, MA, USA) (Current Protocols in Molecular Biology, Ibid). The DNA as subjected to electrophoresis on a 1.0 % NuSeive (FMC BioProducts, Rockland, ME USA) agarose gel. DNA was visualized by exposure to ethidium bromide and long wave uv irradiation. DNA contained in slices isolated from the agarose gel was purified using the Bio 101 GeneClean Kit protocol (Bio 101 Vista, CA, USA)

Cloning, transformation, expression and purification of the PPI gene was carried out essentially as described in Examples II and III above.

*Assay for PPIase activity*

The assay for PPIase was essentially as described by Fisher (Fischer, G., et.al. (1984) *Biomed. Biochim. Acta* 43:1101-1111). The assay measures the *cis-trans* isomerization of the Ala-Pro bond in the test peptide N-succinyl-Ala-Ala-Pro-Phe-p-nitroanilide (Sigma # S-7388, lot # 84H5805). The assay is coupled with  $\alpha$ -chymotrypsin, where the ability of the protease to cleave the test peptide occurs only when the Ala-Pro bond is in *trans*. The conversion of the test peptide to the *trans* isomer in the assay is followed at 390 nm on a Beckman Model DU-650 spectrophotometer. The data were collected every second with an average scanning of time of 0.5 second. Assays were carried out in 35 mM Hepes, pH 8.0, in a final volume of 400  $\mu$ l, with 10  $\mu$ M  $\alpha$ -chymotrypsin (type 1-5 from bovine Pancreas, Sigma # C-7762, lot 23H7020) and 10 nM PPIase. To initiate the reaction, 10  $\mu$ l of the substrate ( 2 mM N-Succinyl-Ala-Ala-Pro-Phe-p-nitroanilide in DMSO) was added to 390  $\mu$ l of reaction mixture at room temperature.

- 120 -

*PPIase assay in crude bacterial extract.*

A 50 ml culture of *Helicobacter pylori* (strain J99) in Brucella broth was harvested at mid-log phase ( $OD_{600\text{ nm}} \sim 1$ ) and resuspended in lysis buffer with the following protease inhibitors: 1 mM PMSF, and 10  $\mu\text{g/ml}$  of each of aprotinin, leupeptin, pepstatine, TLCK, TPCK, and soybean trypsin inhibitor. The suspension was subjected to 3 cycles of freeze-thaw (15 minutes at  $-70^\circ\text{C}$ , then 30 minutes at room temperature), followed by sonication (three 20 second bursts). The lysate was centrifuged (12,000 g x 30 minutes) and the supernatant was assayed for PPIase activity.

**Results**

PPI from *H. pylori* was expressed in *E. coli* using the pET-28b expression vector from Novagen (cat # 69868-1). The expressed recombinant protein was isolated from the soluble fraction of bacterial cells that had been disrupted by cavitation in a Microfluidics Cell disruption chamber. The expression levels of recombinant PPI produced 100 mg of protein. The recombinant protein could be purified to homogeneity by  $\text{Ni}^{2+}$  chelate chromatography and gel filtration. On sodium dodecyl sulfate polyacrylamide gels, the recombinant protein migrates as a single band at 21 kDa, in accordance with the predicted molecular weight of 20,975 deduced from the gene sequence.

The PPIase activity was assayed using the chromogenic tetrapeptide substrate succinyl-Ala-Ala-Pro-Phe-p-nitroanilide. An initial velocity of 4.9  $\mu\text{mole/min/mg}$  protein was measured with the purified enzyme (Figure 5). This corresponds to a  $k_{\text{cat}}$  of  $1.6\text{ sec}^{-1}$  which is similar to the one obtained for the *E. coli* PPIase (Liu, J. and Walsh, C.T. (1990) *Proc. Natl. Acad. Sci. USA* 87:4028-4032) and the one from porcine kidney (Fischer, G. (1989) *Nature* 337:476-478).

The recombinant protein has a high catalytic efficiency of  $2.06 \times 10^9\text{ M}^{-1}\text{ s}^{-1}$  when the assay is measured at  $25^\circ\text{C}$ . These values are one to two orders of magnitude higher than that observed for other characterized PPIases. However, in those studies, the ppiase assay was conducted at  $10^\circ\text{C}$ , which may account for the discrepancy. The catalytic efficiency is very close to the  $1 \times 10^8$  to  $1 \times 10^9\text{ M}^{-1}\text{ s}^{-1}$  upper diffusional limit for "kinetically perfect" enzymes (Albery, W.J. and Knowles, J.R. (1976) *Biochemistry* 15:5631-5640) and suggests that by at least one measure, the *H. pylori* PPIase is a highly effective catalyst in the *cis-trans* isomerisation of the Ala-Pro bond in the oligopeptide substrate.

- 121 -

The presence of PPIase was also determined in an *H. pylori* extract. As with the assay for the recombinant protein, PPIase activity was detected, and was dependent on the concentration of extract added (Figure 6).

These results show that PPIase activity can be measured on either *H. pylori* extracts or on the recombinant protein in *E. coli*. The high catalytic efficiency also demonstrates that *H. pylori* enzymes, such as PPIase, can be expressed at high levels and in an active form in *E. coli*. Such high yields of purified proteins provide for the design of various high throughput drug screening assays.

IX. Cloning, purification, and characterization of the gene encoding the glutamate racemase of *H. pylori*.

The *Helicobacter pylori* genome contains an open reading frame (ORF) of 255 amino acids that was found to have homology to the *Staphylococcus haemolyticus* glutamate racemase gene (*dga*) (NCBI Accession number U12405) and to the *E. coli* *murI* gene which encodes glutamate racemase activity in that organism. To evaluate whether this *H. pylori* ORF encodes a protein with glutamate racemase activity, the gene was isolated by polymerase chain reaction (PCR) amplification cloning, overexpressed in *E. coli*, and the protein purified to apparent homogeneity. A simple assay for glutamate racemase activity resulting in the isomerization of D-glutamic acid to L-glutamic acid was developed to facilitate purification and for future use as a high-throughput drug screen.

The ORF in *H. pylori* has been found by gene disruption studies to be essential for viability of *H. pylori* cells in laboratory culture (see Example VII above). Therefore, inhibition of the enzymatic activity would be expected to be lethal for the organism, and such inhibitors may have utility in antimicrobial therapy of human infectious diseases.

*Cloning of H. pylori murI gene encoding glutamate racemase*

A 765 base pair DNA sequence encoding the *murI* gene of *H. pylori* was isolated by polymerase chain reaction (PCR) amplification cloning. A synthetic oligonucleotide primer (5'-AAATAGTCATATGAAAATAGGCGTTTTTG -3' (SEQ ID NO:1289)) encoding an *NdeI* restriction site and the 5' terminus of the *murI* gene and a primer (5'-AGAATTCTATTACAATTTGAGCCATTCT -3' (SEQ ID NO:1290)) encoding an *EcoRI* restriction site and the 3' end of the *murI* gene were used to amplify the *murI* gene of *H. pylori* using genomic DNA prepared from the J99 strain of *H. pylori* as the template DNA for the PCR amplification reactions (Current Protocols in Molecular Biology, John Wiley and Sons, Inc. F. Ausubel et al., editors, 1994). To amplify a DNA sequence containing the *murI* gene, genomic DNA (25 nanograms) was introduced into

- 122 -

each of two reaction vials containing 1.0 micromole of each synthetic oligonucleotide primer, 2.0 mM MgCl<sub>2</sub>, 0.2 mM of each deoxynucleotide triphosphate (dATP, dGTP, dCTP & dTTP), and 1.25 units of heat stable DNA polymerases (Amplitaq, Roche Molecular Systems, Inc., Branchburg, NJ, USA) in a final volume of 50 microliters.

- 5 The following thermal cycling conditions were used to obtain amplified DNA products for the *murI* gene using a Perkin Elmer Cetus/ GeneAmp PCR System 9600 thermal cycler:

*Conditions for amplification of H. pylori murI;*

- 10 Denaturation at 94°C for 2 min,  
2 cycles at 94°C for 15 sec, 30°C for 30 sec and 72°C for 15 sec  
23 cycles at 94°C for 15 sec, 53°C for 30 sec and 72°C for 15 sec  
Reactions were concluded at 72°C for 20 minutes

- 15 Upon completion of thermal cycling reactions, the amplified DNA was washed and purified using the Qiaquick Spin PCR purification kit (Qiagen, Gaithersburg, MD, USA). The amplified DNA sample was subjected to digestion with the restriction endonucleases, *NdeI* and *EcoRI* (New England Biolabs, Beverly, MA USA) (Current  
20 Protocols in Molecular Biology, Ibid). The DNA samples from each of two reaction mixtures were pooled and subjected to electrophoresis on a 1.0% SeaPlaque (FMC BioProducts, Rockland, ME, USA) agarose gel. DNA was visualized by exposure to ethidium bromide and long wave uv irradiation. Amplified DNA encoding the *H. pylori murI* gene was isolated from agarose gel slices and purified using the Bio 101  
25 GeneClean Kit protocol (Bio 101 Vista, CA, USA).

*Cloning of H. pylori DNA sequences into the pET-23 prokaryotic expression vector.*

- The pET-23b vector can be propagated in any *E. coli* K-12 strain, e.g., HMS174,  
30 HB101, JM109, DH5α, etc., for the purpose of cloning or plasmid preparation. Hosts for expression include *E. coli* strains containing a chromosomal copy of the gene for T7 RNA polymerase. These hosts are lysogens of bacteriophage DE3, a lambda derivative that carries the *lacI* gene, the lacUV5 promoter and the gene for T7 RNA polymerase. T7 RNA polymerase is induced by addition of isopropyl-B-D-thiogalactoside (IPTG),  
35 and the T7 RNA polymerase transcribes any target plasmid, such as pET-28b, carrying its gene of interest. Strains used in our laboratory include: BL21(DE3) (Studier, F.W., Rosenberg, A.H., Dunn, J.J., and Dubendorff, J.W. (1990) Meth. Enzymol. 185, 60-89).

The pET-23b vector (Novagen, Inc., Madison, WI, USA) was prepared for cloning by digestion with *NdeI* and *EcoRI* (Current Protocols in Molecular Biology,

- 123 -

Ibid). Following digestion, the amplified, agarose gel-purified DNA fragment carrying the *murI* gene was cloned (Current Protocols in Molecular Biology, Ibid) into the previously digested pET-23b expression vector. Products of the ligation reaction were then used to transform the BL21(DE3) strain of *E. coli*.

5

*Transformation of competent bacteria with recombinant plasmids*

Competent bacteria, *E. coli* strain BL21 or *E. coli* strain BL21(DE3), were transformed with recombinant pET23-*murI* expression plasmid carrying the cloned *H. pylori* sequence according to standard methods (Current Protocols in Molecular, Ibid).

10 Briefly, 1 microliter of ligation reaction was mixed with 50 microliters of electrocompetent cells and subjected to a high voltage pulse, after which, samples were incubated in 0.45 milliliters SOC medium (0.5% yeast extract, 2.0 % tryptone, 10 mM NaCl, 2.5 mM KCl, 10 mM MgCl<sub>2</sub>, 10 mM MgSO<sub>4</sub> and 20 mM glucose) at 37°C with shaking for 1 hour. Samples were then spread on LB agar plates containing 100  
15 microgram/ml ampicillin for growth overnight. Transformed colonies of BL21 were then picked and analyzed to evaluate cloned inserts as described below.

*Identification of recombinant pET expression plasmids carrying H. pylori sequences*

20 Individual BL21 clones transformed with recombinant pET-23-*murI* were analyzed by PCR amplification of the cloned inserts using the same forward and reverse primers, specific for each *H. pylori* sequence, that were used in the original PCR amplification cloning reactions. Successful amplification verified the integration of the *H. pylori* sequences in the expression vector (Current Protocols in Molecular Biology,  
25 Ibid).

*Isolation and Preparation of plasmid DNA from BL21 transformants*

Colonies carrying pET-23-*murI* vectors were picked and incubated in 5 mls of LB broth plus 100 microgram/ml ampicillin overnight. The following day plasmid  
30 DNA was isolated and purified using the Qiagen plasmid purification protocol (Qiagen Inc., Chatsworth, CA, USA).

*Cloning and expression of the E. coli groE operon*

It has been demonstrated that coexpression of the *E. coli murI* gene with the  
35 genes in the *E. coli groE* operon reduces the formation of insoluble inclusion bodies containing recombinant glutamate racemase (Ashiuchi, M., Yoshimura, T., Kitamura, T., Kawata, Y., Nagai, J., Gorlatov, S., Esaki, N. and Soda, K., 1995, *J. Biochem.* 117,

495-498). The *groE* operon encodes two proteins, GroES (97 amino acids) and GroEL (548 amino acids), which are molecular chaperones. Molecular chaperones cooperate to assist the folding of new polypeptide chains (F. Ulrich Hartl, 1996, *Nature London* 381, pp. 571-580).

5       The 2210 bp DNA sequence encoding the *groE* operon of *E. coli* (NCBI Accession number X07850) was isolated by polymerase chain reaction (PCR) amplification cloning. A synthetic oligonucleotide primer (5'-GCGAATTCGATCAG AATTTTTTTTCT-3' (SEQ ID NO:1291)) encoding an *EcoRI* restriction site and the 5' terminus of the *groE* operon containing the endogenous promoter region of the *groE*  
10       operon and a primer (5'-ATAAGTACTTGTGAATCTTATACTAG -3' (SEQ ID NO:1292)) encoding a *ScaI* restriction site and the 3' end of the *groEL* gene contained in the *groE* operon were used to amplify the *groE* operon of *E. coli* using genomic DNA prepared from *E. coli* strain MG1655 as the template DNA for the PCR amplification reactions (Current Protocols in Molecular Biology, Ibid). To amplify a DNA sequence  
15       containing the *E. coli groE* operon, genomic DNA (12.5 nanograms) was introduced into each of two reaction vials containing 0.5 micromoles of each synthetic oligonucleotide primer, 1.5 mM MgCl<sub>2</sub>, 0.2 mM each deoxynucleotide triphosphate (dATP, dGTP, dCTP & dTTP) and 2.6 units heat stable DNA polymerases (Expanded High Fidelity PCR System, Boehringer Mannheim, Indianapolis, Indiana) in a final  
20       volume of 50 microliters. The following thermal cycling conditions were used to obtain amplified DNA products for the *groE* operon using a Perkin Elmer Cetus/ GeneAmp PCR System 9600 thermal cycler:

25       *Conditions for amplification and cloning of the E. coli groE operon:*

30       Denaturation at 94°C for 2 min,  
2 cycles at 94°C for 15 sec, 30°C for 30 sec and 72°C for 2 min  
23 cycles at 94°C for 15 sec, 55°C for 30 sec and 72°C for 2 min  
Reactions were concluded at 72°C for 8 minutes

      Upon completion of thermal cycling reactions, the amplified DNA was washed and purified using the Qiaquick Spin PCR purification kit (Qiagen, Gaithersburg, MD, USA). The amplified DNA sample was subjected to digestion with the restriction endonucleases, *EcoRI* and *ScaI* (New England Biolabs, Beverly, MA USA) (Current  
35       Protocols in Molecular Biology, Ibid). The DNAs from each of two reaction mixtures were pooled and subjected to electrophoresis in a 1.0% SeaPlaque (FMC BioProducts, Rockland, ME, USA) agarose gel. DNA was visualized by exposure to ethidium bromide and long wave uv irradiation. DNA contained in slices isolated from the

- 125 -

agarose gel was purified using the Bio 101 GeneClean Kit protocol (Bio 101 Vista, CA, USA).

A DNA fragment, *EcoRI* to *Scal*, containing the *E. coli groE* operon was cloned into the corresponding sites of the pACYC184 expression vector (New England Biolabs, Beverly, MA, USA) to make pACYC184-*groE*. The BL21(DE3) strain of *E. coli* was transformed with pACYC-*groE*. A tetracycline-resistant transformant overexpressing proteins of  $M_r \sim 14,000$  (GroES) and  $M_r \sim 60,000$  (GroEL) was isolated.

10 *Transformation of E. coli strain BL21(DE3) carrying the pACYC-groE plasmid of E. coli.*

Competent bacteria derived from a clone of strain BL21(DE3) carrying the pACYC-*groE* plasmid were transformed with 50 nanograms of pET23-*murI* plasmid DNA, isolated as described above (Current Protocols in Molecular Biology, Ibid). A clone of BL21(DE3) carrying both the pACYC-*groE* expression plasmid and the pET-23-*murI* plasmid was isolated and used for expression of recombinant glutamate racemase as described below.

*Expression of recombinant H. pylori murI*

A bacterial clone of BL21(DE3) carrying both the pACYC-*groE* expression plasmid and the pET-23-*murI* plasmid was cultured in LB broth supplemented with 1.0 mM D,L-glutamic acid and 100 microgram/ml ampicillin and 10 micrograms/ml tetracycline at 30°C until an optical density at 600 nM of 0.5 to 1.0 O.D. units was reached, at which point, isopropyl-beta-D-thiogalactoside (IPTG) was added to the culture at a final concentration of 1.0 mM. Cells were cultured overnight to induce gene expression of the *H. pylori* recombinant DNA constructions .

After induction of gene expression with IPTG, bacteria were pelleted by centrifugation in a Sorvall RC-3B centrifuge at 3000 x g for 20 minutes at 4°C. Pellets were resuspended in 50 milliliters of cold 10 mM Tris-HCl, pH 8.0, 0.1 M NaCl and 0.1 mM EDTA (STE buffer). Cells were then centrifuged at 2000 x g for 20 min at 4°C. Pellets were weighed (average wet weight = 6 grams/liter) and processed to purify recombinant protein as described below.

*Purification of soluble glutamate racemase*

All steps were carried out at 4°C. Cells were suspended in 4 volumes of lysis buffer (50 mM Potassium phosphate, pH 7.0, 100 mM NaCl, 2 mM EDTA, 2 mM EGTA, 10% glycerol, 10 mM D,L-glutamic acid, 0.1 %  $\beta$ -mercaptoethanol, 200  $\mu$ g/ml lysozyme, 1 mM PMSF, and 10  $\mu$ g/ml each of leupeptin, aprotinin, pepstatin, L-1-

- 126 -

chloro-3-[4-tosylamido]-7-amino-2-heptanone (TLCK), L-1-chloro-3-[4-tosylamido]-4-phenyl-2-butanone (TPCK), and soybean trypsin inhibitor, and ruptured by three passages through a small volume microfluidizer (Model M-110S, Microfluidics International Corporation, Newton, MA). The resultant homogenate was diluted with 1  
5 volume of buffer A (10 mM Tris-HCl pH 7.0, 0.1 mM EGTA, 10 % glycerol, 1 mM DL-Glutamic acid, 1 mM PMSF, 0.1% beta-mercaptoethanol), made 0.1 % Brij-35, and centrifuged (100,000 x g, 1 h) to yield a clear supernatant (crude extract).

After filtration through a 0.80-um filter, the extract was loaded directly onto a 20 ml Q-Sepharose column pre-equilibrated in buffer A containing 100 mM NaCl and 0.02  
10 % Brij-35. The column was washed with 100 ml (5 bed volumes) of Buffer A containing 100 mM NaCl and 0.02 % Brij-35, then developed with a 100-ml linear gradient of increasing NaCl (from 100 to 500 mM) in Buffer A. A band of  $M_r = 28,000$  corresponding to glutamate racemase, the product of the recombinant *H. pylori murl* gene, eluted at a gradient concentration of approximately 200-280 mM NaCl. Individual  
15 column fractions were then characterized for glutamate racemase activity (see below for description of assay) and the protein profile of the fractions were analyzed on 12 % acrylamide SDS-PAGE gels.

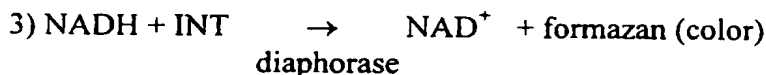
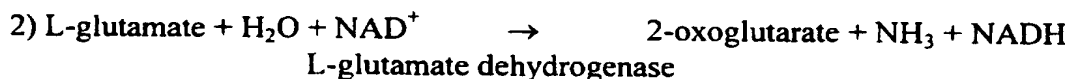
Fractions containing glutamate racemase were pooled, brought to 70% saturation with solid  $(\text{NH}_4)_2\text{SO}_4$ , stirred for 20 min, and then centrifuged at 27,000 x g for 20 min.  
20 The resulting pellet was resuspended in lysis buffer to a final volume of 8 ml and loaded directly onto a 350-ml column (2.2 x 92 cm) of Sephacryl S-100HR gel filtration medium equilibrated in buffer B (10 mM Hepes pH 7.5, 150 mM NaCl, 0.1 mM EGTA, 10% glycerol, 1 mM D,L-glutamic acid, 0.1 mM PMSF, 0.1 % beta-mercaptoethanol) and run at 30 ml/h. Fractions found to contain a glutamate racemase activity were  
25 pooled, and 0.5 volume of buffer C (10 mM Tris pH 7.5, 0.1 mM EGTA, 10% glycerol, 1 mM D,L-glutamic acid, 0.1 mM PMSF, 0.1 % B-mercaptoethanol) was added (to reduce the NaCl concentration to 100 mM), and loaded onto a MonoQ 10/10 high-pressure liquid chromatography column equilibrated in buffer C containing 100 mM NaCl. The column was washed with 5 bed volumes of this buffer and developed with a  
30 40 ml linear gradient of increasing NaCl (from 100 to 500). Glutamate racemase eluted as a sharp peak at 310 mM NaCl. Fractions containing a glutamate racemase activity were pooled, concentrated by dialysis against storage buffer [50% glycerol, 10 mM 3-(N-morpholino-propanesulfonic acid (MOPS) pH 7.0, 150 mM NaCl, 0.1 mM EGTA, 0.02 % Brij-35, 1 mM dithiothreitol (DTT)], and stored at -20°C.

35

- 127 -

*Assays for glutamate racemase activity.*Conversion of D-glutamate to L-glutamate (two enzyme coupled assay)

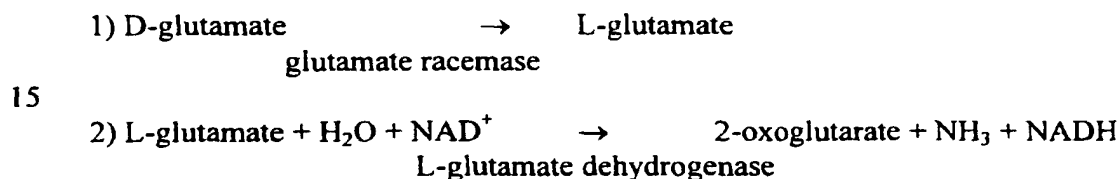
The activity of glutamate racemase, interconversion of the enantiomers of glutamic acid, was measured using D-glutamic acid as substrate. The method of Gallo and Knowles (Gallo, K.A. and Knowles, J.R., 1993, Biochemistry 32, 3981-3990) that was used to measure the glutamate racemase activity of *Lactobacillus fermenti* was adapted for the measurement of glutamate racemase activity of the *H. pylori murI* gene product isolated as a recombinant protein from *E. coli*. In this assay, the measurement of the activity of glutamate racemase is linked to an OD change in the visible range in a series of coupled reactions to the activities of L-glutamate dehydrogenase (reduction of NAD to NADH) and diaphorase (reduction of the dye p-iodonitrotetrazolium violet, INT). Initial rates were determined by following the increase in absorbance at 500 nm in a reaction volume of 200  $\mu$ l containing 50 mM Tris-HCl, pH 7.8, 4% v/v glycerol, 10 mM NAD, 2 mM INT, 60 Units/ml L-glutamate dehydrogenase, 5 Units/ml diaphorase, and varying concentrations of either substrate (from 0.063 mM to 250 mM D-glutamic acid) or purified enzyme (from 1  $\mu$ g to 50  $\mu$ g). After a preincubation of all reagents except either the substrate (D-glutamic acid) or the enzyme (*murI* gene product) for a period of 5 minutes, reactions were initiated by adding the missing ingredient (i.e., the enzyme or the substrate, as required), and the increase in optical density at 500 nm was measured in a Microplate Spectrophotometer System (Molecular Devices, Spectra MAX 250). Measurements were followed for 20 minutes, and initial velocities were derived by calculating the maximum slope for the absorbance increases. The coupled reaction can be summarized as shown below:

Conversion of D-glutamate to L-glutamate (single enzyme coupled assay)

In this assay, the conversion of D-glutamic acid to L-glutamic acid is coupled to the conversion of L-glutamic acid and  $\text{NAD}^+$  by L-glutamate dehydrogenase to 2-oxoglutarate, ammonia. The production of NADH is measured as an increase of

- 128 -

absorbance at 340 nm (the reduction of  $\text{NAD}^+$  to NADH) at 37°C. The standard assay mixture (adapted from Choi, S-Y., Esaki, N., Yoshimura, T., and Soda, K., 1991, Protein Expression and Purification 2, 90-93) contained 10 mM Tris-HCl, pH 7.5, 5 mM  $\text{NAD}^+$ , 5 Units/ml L-glutamate dehydrogenase, varying concentrations of the substrate D-Glutamic Acid (0.063 mM to 250 mM), and the purified recombinant *H. pylori* enzyme glutamate racemase (1  $\mu\text{g}$  to 50  $\mu\text{g}$ ). The reaction was started by the addition of either the substrate D-glutamic acid or the recombinant glutamate racemase after a preincubation at 37°C for 5 minutes with all of the other assay ingredients. The change in absorbance at 340 nm was measured in a Spectra MAX 250. Initial velocities were derived from the initial slopes. The coupled reactions can be summarized as shown below:



## Results

### 1) Expression of the *H. pylori murI* gene in *E. coli* cells

To examine its biochemical properties, the *H. pylori* glutamate racemase was overexpressed in *E. coli* and purified. In the presence of the *E. coli* chaperones GroES and GroEL, the glutamate racemase was expressed as a soluble protein. About 20 mg of soluble Murl was produced per liter of culture as judged by intensity of the protein band after SDS-PAGE. No band corresponding to the molecular weight of *murI* protein was seen in control gel lanes containing extracts from cells transformed with the pET vector lacking a *murI* insert. Addition of 1 mM DL-glutamic acid during cultivation of the expressing cells increased the apparent expression level by about five-fold.

### 2) Purification of recombinant *H. pylori murI* protein

*MurI* was purified by cation exchange chromatography and gel filtration. Upon SDS-PAGE analysis, the purified protein migrated as a single polypeptide species with an apparent mass 29 kDa which is consistent with the predicted mass of 28.858.

- 129 -

### 3) Kinetic properties of recombinant *H. pylori* *murI* enzyme

Kinetic constant for recombinant glutamate racemase were estimated by assaying its activity at various concentrations of protein and D-glutamic acid as described above. Purified recombinant *H. pylori* glutamate racemase exhibits a  $V_{\max}$  of ~ 300 nmoles/min/mg protein ( $k_{\text{cat}} = 8.6 \text{ min}^{-1}$ ) and a  $K_m$  of ~100  $\mu\text{M}$  for D-glutamate. Although the  $V_{\max}$  value is lower than that observed for highly purified glutamate racemase from some other bacterial species, its  $K_m$  for D-glutamic acid is higher than that observed for the enzyme from most other species, resulting in a catalytic efficiency ( $k_{\text{cat}}/K_m$ ) which is typical of purified preparation from *E. coli* and *P. Pentococcus*.

### 4) Characterization of *MurI*: Inhibition by L-serine-O sulfate

The *H. pylori* glutamate racemase was tested for inactivation with a suicide inhibitor, L-serine-O sulfate, which is known to inhibit *murI* from *E. coli*. The enzyme was incubated in the presence of 20 mM L-serine-O sulfate, and at different times interval, aliquots were removed to determine residual activity. The initial velocity of purified recombinant *H. pylori* *murI* protein was determined in the single enzyme coupled assay following incubation with the inhibitor L-serine-O-sulfate (LSOS) at 20 mM for the times indicated on the x-axis. The control was incubated in an identical manner but without LSOS. As shown in Figure 7, the *H. pylori* glutamate racemase can be readily inactivated by the inhibitor.

### Future application of the glutamate racemase activity in high throughput drug screening assays.

The assays for measurement of *H. pylori* glutamate racemase activity described above have been carried out in 96-well plates in which multiple reactions were conducted simultaneously. Measurements of activity in a multi-well format are readily amenable to scale-up to permit rapid analysis of numerous compounds for inhibition of the glutamate racemase activity. Compounds which inhibit the activity of glutamate racemase may have application as novel antibiotics and may be suitable for the treatment and eradication of bacterial (e.g., *H. pylori*) infections in humans. Known inhibitors of glutamate racemase, such as L-serine-O-sulfate, can be used to calibrate high throughput screens of new compound libraries to facilitate identification of new compounds with properties suitable for *in vivo* human therapeutics.

- 130 -

**EQUIVALENTS**

Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments and methods described herein. Such equivalents are intended to be encompassed by the scope of the

5 following claims.

131

## SEQUENCE LISTING

## (1) GENERAL INFORMATION:

## (i) APPLICANT:

- (A) NAME: Astra Aktiebolag
- (B) STREET: S-151 85
- (C) CITY: Sodertalje
- (D) STATE:
- (E) COUNTRY: Sweden
- (F) POSTAL CODE (ZIP)

- (ii) TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES  
RELATING TO HELICOBACTER PYLORI AND  
VACCINE COMPOSITIONS THEREOF

- (iii) NUMBER OF SEQUENCES: 1298

## (iv) CORRESPONDENCE ADDRESS:

- (A) ADDRESSEE: LAHIVE & COCKFIELD
- (B) STREET: 60 State Street, Suite 510
- (C) CITY: Boston
- (D) STATE: Massachusetts
- (E) COUNTRY: USA
- (F) ZIP: 02109-1875

## (v) COMPUTER READABLE FORM:

- (A) MEDIUM TYPE: CD-ROM ISO9660
- (B) COMPUTER:
- (C) OPERATING SYSTEM:
- (D) SOFTWARE:

## (vi) CURRENT APPLICATION DATA:

- (A) APPLICATION NUMBER:
- (B) FILING DATE:

## (vii) PRIOR APPLICATION DATA:

- (A) APPLICATION NUMBER: US 08/625,811
- (B) FILING DATE: 29-MAR-1996

## (viii) PRIOR APPLICATION DATA:

- (A) APPLICATION NUMBER: US 08/758,731
- (B) FILING DATE: 02-APR-1996

## (ix) PRIOR APPLICATION DATA:

- (A) APPLICATION NUMBER: US 08/736,905
- (B) FILING DATE: 25-OCT-1996

## (x) PRIOR APPLICATION DATA:

- (A) APPLICATION NUMBER: US 08/738,859
- (B) FILING DATE: 28-OCT-1996

(xi) PRIOR APPLICATION DATA:

- (A) APPLICATION NUMBER: US 08/761,318
- (B) FILING DATE: 06-DEC-1996

(xii) ATTORNEY/AGENT INFORMATION:

- (A) NAME: Mandragouras, Amy E.
- (B) REGISTRATION NUMBER: 36,207
- (C) REFERENCE/DOCKET NUMBER: GTN-009C2PC

(xiii) TELECOMMUNICATION INFORMATION:

- (A) TELEPHONE: (617)227-7400
- (B) TELEFAX: (617)227-5941

268

```

TAAAAAAAG CGTTTAAGGA GAGGTTTTGC TCTCAAGTGT ATATCTCTTT TAATGTGGAT 120
CACAACTCTT TATCCGCTCA AGTTTTAAGG GTTAAAAACC ACCGCATCAA AGAGAAATTT 180
TTTAAAACTT TTGAGACTAA AGTGGAAGCT AAAAATGGTG AAGTCCCAT TCAAGCCTTA 240
AAAATCGCCA GAACTTATAG CCAAAAATAC CCCTACACCT ATTTTAGCGC GATGAGTAAA 300
GCTAAAGAGG TTTTATGCGA AAAGCAGGCG TTTGAACAAA TCAAACAAGA AAATCAAGAT 360
TATCAGGCTT GTGAAGTCAA TCAAAAGTAT TGCCTTTATG TGGAATCTAA GGATTTTCTA 420
AAGGATTTCA AGCGTTTTAA AATCCAGGAT GTGGATTTTT TGTTTTCGCC CTTTAGCCTT 480
ATTTATGATT TTGTGCGCGA TAATTTAGAA AACAAGCCGT TGTGTATCT GCTTTTAGAG 540
CGTTCAAGGT TTTATTTTTT GATTGCGGAT AAAAAAGAGA TTTTTTATAG CAAATCCGTG 600
TTTTTAGAAG AGCAACCTGA AGAGTTTATA GAGAGCAAAG AAGAAGATTC TATGGGAATG 660
GATAATGAAG CTGTGGATTT GTTTTTGAGT GAAATCCAAG AAGATATTGA CAGCCTTGAA 720
GAAGCGATAG GCCTAGACAG CAGCAAGGAT AATAGCGAAA AAATAACAGA GGACGCTTAT 780
AGTTTGATTG AAGGCATGAC GAATATCCCC TTGATTGCAG ATGTTTTGCA AGAGGGATTG 840
CGTGCGCTCT ATCATTCTAG AGAGATAGAC TTTGTAGAAA AAGTGGTTGT TTTAGACAGC 900
TGTCAAATCC ACCACAAGGC GTTAATGCAT TTGCAAGAAA CTTTGATGAT AGAAGTGGAT 960
AGGCTCGATT TTTCTTTAGT GGAGCGGTTG AATGTTTTAG CGCGCATGGA GAATGAAAAG 1020
CATGCGTTT 1029

```

## (2) INFORMATION FOR SEQ ID NO:227:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1260 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: circular

## (ii) MOLECULE TYPE: DNA (genomic)

## (iii) HYPOTHETICAL: NO

## (iv) ANTI-SENSE: NO

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: *Helicobacter pylori*

## (ix) FEATURE:

- (A) NAME/KEY: misc\_feature
- (B) LOCATION 1...1260

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:227:

```

ATGATTTTTG GGGATTTTAA ATATCAAAAA AGCGTTAAAA AACTCACAGC CACTAACCTT 60
AACGAGCTTA AAAACGCCCT GGATTTTATC TCTCAAAATA GGGGAAGGG GTATTTTGTG 120
GGGTATCTTT TGTATGAAGC GCGCTTAGCG TTTTTAGATG AAAATTTTCA AAGCCAAACC 180
CCTTTTTTGT ATTTTGAACA ATTTTTAGAA AGAAAAAAT ACCTTGTA GAATCTAAAA 240
GAGCATGCGT TTTACCCTAA AATCCACAGC TCTTTAGATC AAAAAACTTA TTTCAAGCAG 300
TTTAAGGCCG TTAAAGAGCA TCTCAAAAAC GGCGATACCT ATCAAGTGAA TTTGACAATG 360
GATTTATTGT TAAACACTAA GGCCAAACCA AAGCGCGTTT TTAAGGAAGT GATACACAAC 420
CAAAACACGC CTTTTAAGGC TTTTATAGAA AATGAGTTTG GGAGCGTTTT AAGCTTTTCG 480
CCGAATTGT TTTTGAATT AGAGTTTTTA GACACAGCGA TTAAGATTAT CACAAAACCC 540
ATGAAAGGCA CGATCGCTCG CTCAAACAAC CCCTTGATAG ATGAAAAAAA CCGATTGTTT 600
TTGCAAAATG ATGACAAAAA TAGAAGCGAA AATGTGATGA TCGTGGATTT ATTGCGTAAC 660
GATTTGAGCC GCTTGCCCTT AAAAAATAGC GTGAAAGTCA ATCAATTGTT TGAATCATC 720
AGTTTGCCGA GCGTGTATCA AATGATAAGC GAGATTGAAG CTCAATTGCC CCTAAAACA 780
AGCTTGTTTG AGATTTTAA GGCGTTGTTT CCTTGCGGCT CTGTGACCGG ATGCCCCAAA 840
ATCAAAACCA TGCAATCAT TGAAAGTTTA GAAAAACGCC CTAGGGGGGT GTATTGTGGG 900
CGGATAGGCA TGCTTGAGG AAAAAAGCC CTTTTTAGCG TGCCTATCCG CACTTTAGAA 960

```

269

AAAAGAGCGC	ATGAAGATTT	TTTGCATTTA	GGGGTAGGGA	GTGGGGTAAC	TTATAAAAAGT	1020
AAAGCTTCAA	AGGAATATGA	AGAGAGCTTT	TTAAAATCCT	TTTTTGTGAT	GCCCCAAAATA	1080
GAATTTGAGA	TCGTGGAAAC	GATGAGAGTT	ATCAAAAAGG	ATCAAAAATT	AGAGATTAAAC	1140
AATAAAAACG	CCCATAAAGA	ACGCTTAATG	CATAGCGCCC	AATATTTTAA	CTTTAATACA	1200
GATGAAAATC	TTTtagactt	TGAATTAGAA	GGAGAAGGGG	TTTtaAGGGT	TTTAATTCAA	1260

## (2) INFORMATION FOR SEQ ID NO:228:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 855 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: circular

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: *Helicobacter pylori*

(ix) FEATURE:

- (A) NAME/KEY: misc\_feature
- (B) LOCATION 1...855

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:228:

AAACAAATAA	GAAAAGAACG	CTTGAACAAA	CTGCTTAAAA	AGGGGTTTTT	AGCGTTCTTT	60
TTAAGCGTGT	ATTTAAGGGC	TGATGATCTG	GTTACTTATA	CGATCGCCAA	AGAAGAAGAT	120
CTAGGGTACC	AGCGGTTTTT	AGCCAAGAAG	TGTTTAAGGG	GTAAAACCCA	CCCCCATGT	180
TTTACCGCGT	CTAAAAAGCC	CAAGAGGAAA	CCTTTTAATA	TAGATAAAAG	CTCCCATTAT	240
TATGGCACGA	GCGTGGTGCA	AATGTCATGG	CTACAGAGTA	GGGAAAAATT	TGAAAACCAT	300
TCAAAATACC	GAAACATTCC	TTTTGCTGAA	GTCAGTTTGA	TTTATGGCTA	TAAACAATTT	360
TTTCCTAAAA	AAGAGCACTA	CGGCTTCCGT	TTTTATGTCT	CTTTGGATTA	CGCTTATGGC	420
TTTTTTCTTA	AAAATAAAGG	CGCATTAGGC	GATAGTTTGA	GGGCGAGTTC	GCAGATCCCT	480
AAAAGCTATA	GAGAAAAATT	ACAAAGAAAA	GAGACTTTTA	TTAACGCTAT	TTTTTATGGC	540
GTGGGGGCTG	ACTTTTTATA	CAAACGCGCT	TTTGGGACGC	TGATTTTAGG	GGTGAATTTT	600
GTGGGAGAAA	CCTGGTTTTA	TGAAACAAAG	ATTTTCAAAC	AATGGGCTAA	AGACTCTCTA	660
AATACTTACC	GCCCCAACAT	GTTTCAAGTG	ATGTTGAATG	TGGGGTATCG	CTACCGCTTT	720
TCAAGGTATA	AGAATTGGGC	GATAGAATTT	GGCGCGCGCA	TCCCCTTTTT	AATTAATGAT	780
TATTTTAAAA	CCCCTTTATA	CACCCTTCAT	TTCAAACGCA	ATATTTCTGT	CTATCTCACT	840
TCAACTTATG	ATTTT					855

## (2) INFORMATION FOR SEQ ID NO:229:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 693 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: circular

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

344

GTCAAATACG	AAGATAAGGG	CGAACCTCTC	AAAATCACAT	GGCATTCTAA	TTGCCATGCC	420
TTAAGGGTGG	CTAAAGTGAT	TGACTCGGCG	AAAAATCTCA	TCAGACAGCT	TAAAAATGTG	480
GAACTCATTG	AATTGGAAAA	AGAAGAAGAA	TGCTGCGGGT	TTGGGGGGAC	TTTTTCAGTT	540
AAAGAGCCTG	AAATTTTCAGC	GGTTATGGTC	AAAGAAAAGA	TTAAAGACAT	AGAGAGCCGT	600
CATGTGGATG	TGATCGTTTC	AGCGGATGCT	GGGTGCTTGA	TGAATATCAG	CACCGCTATG	660
CAAAAAATGG	GCTCTTTGAC	AAAACCCATG	CATTTTTTATG	ACTTTTTTAGC	CTCAAGGCTT	720
GGACTTTAA						729

## (2) INFORMATION FOR SEQ ID NO:345:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1206 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: circular

## (ii) MOLECULE TYPE: DNA (genomic)

## (iii) HYPOTHETICAL: NO

## (iv) ANTI-SENSE: NO

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: *Helicobacter pylori*

## (ix) FEATURE:

- (A) NAME/KEY: misc\_feature
- (B) LOCATION 1...1206

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:345:

ATGATTTTTG	GGGATTTTAA	ATATCAAAAA	AGCGTTAAAA	AACTCACAGC	CACTAACCTT	60
AACGAGCTTA	AAAACGCCCT	GGATTTTATC	TCTCAAAATA	GGGGGAAGGG	GTATTTTGTG	120
GGGTATCTTT	TGTATGAAGC	GCGCTTAGCG	TTTTTAGATG	AAAATTTTCA	AAGCCAAACC	180
CCTTTTGTGT	ATTTTGAACA	ATTTTATAGAA	AGAAAAAAAT	ACCCTTTAGA	GCCTTTAAAA	240
GAGCATGCGT	TTTACCCTAA	AATCCACAGC	TCTTTAGATC	AAAAAACTTA	TTTCAAGCAG	300
TTTAAGGCCG	TTAAAGAGCA	TCTCAAAAAC	GGCGATACCT	ATCAAGTGAA	TTTGACAATG	360
GATTTATTGT	TAAACACTAA	AGCCAAACCA	AAGCGCGTTT	TTAAGGAAGT	GATACACAAC	420
CAAAACACGC	CTTTTAAGGC	TTTTATAGAA	AATGAGTTTG	GGAGCGTTTT	AAGCTTTTCG	480
CCGGAATTGT	TTTTTGAATT	AGAGTTTTTA	GACACAGCGA	TTAAGATTAT	CACAAAACCC	540
ATGAAAGGCA	CGATCGCTCG	CTCAAACAAC	CCCTTGATAG	ATGAAAAAAA	CCGATTGTTT	600
TTGCAAAATG	ATGACAAAAA	TAGAAGCGAA	AATGTGATGA	TCGTGGATTT	ATTGCGTAAC	660
GATTTGAGCC	GCTTGCCCTT	AAAAAATAGC	GTGAAAGTCA	ATCAATTGTT	TGAAATCATC	720
AGTTTGCCGA	GCGTGTATCA	AATGATAAGC	GAGATTGAAG	CTCAATTGCC	CCTAAAAACA	780
AGCTTGTTTG	AGATTTTAA	GGCGTTGTTT	CCTTGCGGCT	CTGTGACCGG	ATGCCCTAAA	840
ATCAAAACCA	TGCAAATCAT	TGAAAGTTTA	GAAAAACGCC	CTAGGGGGGT	GTATTGTGGG	900
GCGATAGGCA	TGGTTGGAGG	AAAAAAAGCC	CTTTTATAGC	TGCCTATCCG	CACTTTAGAA	960
AAAAGAGCGC	ATGAAGATTT	TTTGCATTTA	GGGGTAGGGA	GTGGGGTAAC	TTATAAAAGT	1020
AAAGCTTCAA	AGGAATATGA	AGAGAGCTTT	TTAAAATCCT	TTTTTGTGAT	GCCCCAAATA	1080
GAATTTGAGA	TCGTGGAAAC	GATGAGAGTT	ATCAAAAAGG	ATCAAAAATT	AGAGATTAAC	1140
AATAAAAACG	CCCATAAAGA	ACGCTTAATG	CATAGCGCCC	AATATTTTAA	CTTTAAATAC	1200
AGATGA						1206

## (2) INFORMATION FOR SEQ ID NO:346:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 795 base pairs

909

GAAAAATAGGG	ATAATGTGGA	AAAACAAGCG	ATTAGAGATC	CTAATATCAA	AGAATTCGCA	720
TGCGGGAAGT	GGGTTTATGA	TGATGAAAAT	TTACAAGCCT	ATCGCCCAAG	CATTTTAAAA	780
CGCGTTGATA	AAGACAAAGA	GATTACAACA	GACATTACCC	CTTGCGATTA	CAGCACCGCT	840
GAAAATAAAA	GCGGTAAAAT	CATTACCCCC	TATACTAAAA	TCTCTGTTCA	TAAAACAGAG	900
CCTTTAGAAG	ACCCTCAAAC	TTTTGAAGCT	AAAAATAATT	TCGCCATTCT	CCAAGCCAGA	960
AGCTCTACAG	AAAAGTGTA	AAGGGCTAGA	GCGAGAAAAG	ACGGCACGAC	CAGGCAATGC	1020
TATCTGATAG	AAGAGCCTTT	AAAACAAGCG	TGGGAGAGCG	AGTATGAAAT	CACCACGCAA	1080
TTAGTGAAAG	CCATATATGA	ACGCCCCAAA	CAAGACGATC	AAGTAGAGCC	GACTTTTAT	1140
GAAACCAGCG	AATTGGCTTA	TTCTTCCACA	CGAAAAAGCG	AAATAACGCG	TAATGAATTG	1200
AATTTGAATG	AAAAATTCAT	GGAATTTGTG	GAAGTGATG	AGGGGCATTA	TTTAAACGAT	1260
ATAATCAAAG	AGAGCAGCGA	ATACAAAGAA	TGGGTAAAAA	ACCATGTGCG	CTTTAAAGAA	1320
GGGGTGTGCA	TGGCCTTAGA	AATAGAAGAG	CAACCACGAG	CCAAAAGCAC	GCCTTTGAGT	1380
ATTGAAAAC	CTCGTGTGT	GTGTGTCAA	AAGGGGAATT	ATTTATTCAA	TGAAGTTTAA	1440

## (2) INFORMATION FOR SEQ ID NO:1003:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1704 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: circular

## (ii) MOLECULE TYPE: DNA (genomic)

## (iii) HYPOTHETICAL: NO

## (iv) ANTI-SENSE: NO

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: *Helicobacter pylori*

## (ix) FEATURE:

- (A) NAME/KEY: misc\_feature
- (B) LOCATION 1...1704

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:1003:

ATGATTTTTG	GGGATTTTAA	ATATCAAAAA	AGCGTTAAAA	AACTCACAGC	CACTAACCTT	60
AACGAGCTTA	AAAACGCCCT	GGATTTTATC	TCTCAAAATA	GGGGGAAGGG	GTATTTTGTG	120
GGGTATCTTT	TGTATGAAGC	GCGCTTAGCG	TTTTTAGATG	AAAATTTTCA	AAGCCAAACC	180
CCTTTTTTGT	ATTTTGAACA	ATTTTATAGAA	AGAAAAAAT	ACCCTTTAGA	GCCTTTAAAA	240
GAGCATGCGT	TTTACCCTAA	AATCCACAGC	TCTTTAGATC	AAAAAACTTA	TTTCAAGCAG	300
TTTAAGGCCG	TTAAAGAGCA	TCTCAAAAAC	GGCGATACCT	ATCAAGTGAA	TTTGACAATG	360
GATTTATTGT	TAAACACTAA	AGCCAAACCA	AAGCGCGTTT	TTAAGGAAGT	GATACACAAC	420
CAAAACACGC	CTTTTAAGGC	TTTTATAGAA	AATGAGTTTG	GGAGCGTTT	AAGCTTTTTCG	480
CCGGAATTGT	TTTTTGAATT	AGAGTTTTTA	GACACAGCGA	TTAAGATTAT	CACAAAACCC	540
ATGAAAGGCA	CGATCGCTCG	CTCAAACAAC	CCCTTGATAG	ATGAAAAAAA	CCGATTGTTT	600
TTGCAAAAATG	ATGACAAAAA	TAGAAGCGAA	AATGTGATGA	TCGTGGATTT	ATTGCGTAAC	660
GATTTGAGCC	GCTTGGCCTT	AAAAAATAGC	GTGAAAGTCA	ATCAATTGTT	TGAAATCATC	720
AGTTTGCCGA	GCGTGTATCA	AATGATAAGC	GAGATTGAAG	CTCAATTGCC	CCTAAAAACA	780
AGCTTGTTTG	AGATTTTTTAA	GGCGTTGTTC	CCTTGCGGCT	CTGTGACCGG	ATGCCCTAAA	840
ATCAAAAACCA	TGCAATCAT	TGAAAGTTTA	GAAAAACGCC	CTAGGGGGGT	GTATTGTGGG	900
GCGATAGGCA	TGGTTGGAGG	AAAAAAAGCC	CTTTTATAGC	TGCCTATCCG	CACTTTAGAA	960
AAAAGAGCGC	ATGAAGATTT	TTTGCATTTA	GGGTTAGGGA	GTGGGGTAAC	TTATAAAAGT	1020
AAAGCTTCAA	AGGAATATGA	AGAGAGCTTT	TTAAAATCCT	TTTTTGTGAT	GCCCCAAATA	1080
GAATTTGAGA	TCGTGGAAAC	GATGAGAGTT	ATCAAAAAGG	ATCAAAAATT	AGAGATTAAC	1140
AATAAAAACG	CCCATAAAGA	ACGCTTAATG	CATAGCGCCC	AATATTTTAA	CTTTAAATAC	1200

910

GATGAAAATC	TTTTAGACTT	TGAATTAGAA	AAAGAAGGGG	TTTAAAGGGT	TTTACTCAAT	1260
AAAAGGGGCA	AGCTCATTA	AGAATACAGA	ACCTTAGAGC	CTTTAAAAAG	CCTAGAAAATC	1320
CGTTTGAGTG	AAGCCCCCAT	TGACAAACAC	GATGATTTTT	TATACCATAA	GACCGCTTAT	1380
GCCCCTTTTT	ATCAAAACGC	TCGAGCGCTC	ATTAAAAAAG	GCGTTATTTT	TGATGAAAATC	1440
TTTTATAACC	AGGATTGGA	ACTCACTGAG	GGCGCTAGGA	GTAATCTTAT	TTTAGAAAATC	1500
CACAACAGGC	TTTTAACCCC	TTATTTTCAGC	GTGGGCGCAT	TAACCGGGAC	GGGCGTTGTG	1560
GGGTTGTTAA	AAAAAGGTCT	TGTTGAGCAT	GCCCCTTTAA	AATTACAAGA	CTTGCAAAGA	1620
GCGGCTAAAA	TCTATTGTAT	TAACGCGCTA	TACGGCTTAG	TGGAGGTGGG	AATAATAGGT	1680
TACCCAATGG	AGCAAAAAAG	TTAA				1704

## (2) INFORMATION FOR SEQ ID NO:1004:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1125 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: circular

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: *Helicobacter pylori*

(ix) FEATURE:

- (A) NAME/KEY: misc\_feature
- (B) LOCATION 1...1125

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1004:

ATGAAAATCA	GTGTTAGTAA	AAACGATTTA	GAAAACACTT	TGCGCTACTT	GCAAGCTTTT	60
TTGGATAAAA	AGGACGCTTC	CTCTATCGCT	TCACACATCC	ATTTAGAAGT	CATTAAAGAA	120
AAGCTCTTTT	TAAAAGCCAG	CGATTTCAGAT	ATTGGACTAA	AAAGCTATAT	TTCTACGCAA	180
TCTACCGATA	AAGAGGGCGT	AGGCACGATC	AACGGGAAGA	AGTTTTTAGA	TATTATCTCA	240
TGTTTGAAAAG	ACTCTAATAT	TGTTTTAGAG	ACTAAAGATG	ACAGCTTGGT	GATCAAACAA	300
AATAAAAGCT	CTTTCAAAC	CCCCATGTTT	GACGCTGATG	AGTTCCCTGA	ATTCCTTGTT	360
ATTGATCCAA	AAGTGAGTTT	AGAAATCAAT	GCCCCCTTTT	TGGTGATGTC	GTTTAAAAAG	420
ATCGCTCCTG	TGATTGAGCA	AACTAGCCAT	AAAAGGGAAT	TAGCCGGTGT	TTTAATGCAA	480
TTCAATCAAA	AACACCAAAC	CCTTTCAGTG	GTAGGCACGG	ATACCAAGCG	GCTCTCTTAC	540
ACGCAGTTAG	AAAAAATCTC	TATCCATTCC	ACCGAAGAAG	ACATCTCTTG	CATTTTGCCT	600
AAAAGGGCTT	TATTGGAAAT	CCTTAAACTT	TTTTATGAAA	ATTTAGTTT	TAAAAGCGAC	660
GGCATGTTAG	CGGTGGTTGA	AAACGAAACG	CACGCTTTTT	TCACCAAGCT	CATTGATGGG	720
AATTACCCTG	ATTATCAAAA	AATCCTCCCT	AAAGAATACA	CTTCTCTTT	CACTTTAGGC	780
AAGGAAGAAT	TTAAAGAGGG	CATTAAGTTG	TGCAGTTCTT	TAAGCTCCAC	CATTAAGCTC	840
ACTTTAGAAA	AAAACAACGC	TTTGTTTGAA	TCTTTGGATT	CTGAGCATAG	CGAAACCGCC	900
AAAACCTCTG	TTGAGATTGA	AAAAGGTTTG	GATATTGAAA	AAGCCTTTCA	TTTGGGCGTG	960
AACGCAAAAT	TTTTCTTGA	AGCCTTAAAC	GCTTTAGGGA	CAACGCAATT	TGTTTTAAAA	1020
TGCAATGAGC	CTTCTTCGCC	TTTTTTGATT	CAAGAACCTC	TTGATGAAAA	GCAAAGCCAC	1080
TTGAACGCTA	AAATTTCCAC	TTTGATGATG	CCAATCACAC	TATAA		1125

## (2) INFORMATION FOR SEQ ID NO:1005:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1107 base pairs

- 1110 -

## CLAIMS

1. An isolated nucleic acid comprising a nucleotide sequence encoding an *H. pylori* polypeptide selected from the group consisting of SEQ ID NO: 492-SEQ ID NO: 759, SEQ ID NO: 761, SEQ ID NO: 763, SEQ ID NO: 765-SEQ ID NO: 818, SEQ ID NO: 820-SEQ ID NO: 846, SEQ ID NO: 848-SEQ ID NO: 896, SEQ ID NO: 898-SEQ ID NO: 963, SEQ ID NO: 966-SEQ ID NO: 982, SEQ ID NO: 1037, SEQ ID NO: 1038, SEQ ID NO: 1041-SEQ ID NO: 1087, SEQ ID NO: 1090 and SEQ ID NO: 1296-SEQ ID NO: 1298.
2. A recombinant expression vector comprising the nucleic acid of claim 1 operably linked to a transcription regulatory element.
3. A cell comprising a recombinant expression vector of claim 2.
4. A method for producing an *H. pylori* polypeptide comprising culturing a cell of claim 3 under conditions that permit expression of the polypeptide.
5. A probe comprising a nucleotide sequence consisting of at least 8 nucleotides of a nucleotide sequence selected from the group consisting of SEQ ID NO: 1-SEQ ID NO: 268, SEQ ID NO: 270, SEQ ID NO: 272, SEQ ID NO: 274-SEQ ID NO: 327, SEQ ID NO: 329-SEQ ID NO: 364, SEQ ID NO: 366-SEQ ID NO: 405, SEQ ID NO: 407-SEQ ID NO: 472, SEQ ID NO: 475-SEQ ID NO: 491, SEQ ID NO: 983, SEQ ID NO: 984, SEQ ID NO: 987-SEQ ID NO: 1033, SEQ ID NO: 1036 and SEQ ID NO: 1293-SEQ ID NO: 1295 or the complement thereof.
6. An isolated nucleic acid comprising a nucleotide sequence of at least 8 nucleotides in length, wherein the sequence is hybridizable to a nucleic acid having a nucleotide sequence selected from the group consisting of SEQ ID NO: 1-SEQ ID NO: 268, SEQ ID NO: 270, SEQ ID NO: 272, SEQ ID NO: 274-SEQ ID NO: 327, SEQ ID NO: 329-SEQ ID NO: 364, SEQ ID NO: 366-SEQ ID NO: 405, SEQ ID NO: 407-SEQ ID NO: 472, SEQ ID NO: 475-SEQ ID NO: 491, SEQ ID NO: 983, SEQ ID NO: 984, SEQ ID NO: 987-SEQ ID NO: 1033, SEQ ID NO: 1036 and SEQ ID NO: 1293-SEQ ID NO: 1295 or the complement thereof.

- 1111 -

7. A vaccine composition for prevention or treatment of an *H. pylori* infection comprising an effective amount of an isolated nucleic acid comprising a nucleotide sequence encoding an *H. pylori* polypeptide or a fragment thereof, said nucleic acid comprising a nucleotide sequence selected from the group consisting of  
5 SEQ ID NO: 1-SEQ ID NO: 268, SEQ ID NO: 270, SEQ ID NO: 272, SEQ ID NO: 274-SEQ ID NO: 327, SEQ ID NO: 329-SEQ ID NO: 364, SEQ ID NO: 366-SEQ ID NO: 405, SEQ ID NO: 407-SEQ ID NO: 472, SEQ ID NO: 475-SEQ ID NO: 491, SEQ ID NO: 983, SEQ ID NO: 984, SEQ ID NO: 987-SEQ ID NO: 1033, SEQ ID NO: 1036 and SEQ ID NO: 1293-SEQ ID NO: 1295.

10

8. A vaccine composition of claim 7, further comprising a pharmaceutically acceptable carrier.

9. A vaccine composition of claim 8, wherein the pharmaceutically  
15 acceptable carrier is an adjuvant.

10. A method of treating a subject for *H. pylori* infection comprising administering to a subject a vaccine composition of claim 7, such that treatment of *H. pylori* infection occurs.

20

11. A method of claim 10, wherein the treatment is a prophylactic treatment.

12. A method of claim 10, wherein the treatment is a therapeutic treatment.

25 13. A method for detecting the presence of a *Helicobacter* nucleic acid in a sample comprising:

(a) contacting a sample with a probe of claim 5 under conditions in which a hybrid can form between the probe and a *Helicobacter* nucleic acid in the sample; and

30 (b) detecting the hybrid formed in step (a), wherein detection of a hybrid indicates the presence of a *Helicobacter* nucleic acid in the sample.

14. A recombinant or substantially pure preparation of an *H. pylori* polypeptide selected from the group consisting of SEQ ID NO: 492-SEQ ID NO: 759,  
35 SEQ ID NO: 761, SEQ ID NO: 763, SEQ ID NO: 765-SEQ ID NO: 818, SEQ ID NO: 820-SEQ ID NO: 846, SEQ ID NO: 848-SEQ ID NO: 896, SEQ ID NO: 898-SEQ ID

- 1112 -

NO: 963, SEQ ID NO: 966-SEQ ID NO: 982, SEQ ID NO: 1037, SEQ ID NO: 1038, SEQ ID NO: 1041-SEQ ID NO: 1087, SEQ ID NO: 1090 and SEQ ID NO: 1296-SEQ ID NO: 1298.

- 5           15.     A vaccine composition for prevention or treatment of an *H. pylori* infection comprising an effective amount of a purified *H. pylori* polypeptide or a fragment thereof, wherein said *H. pylori* polypeptide is selected from the group consisting of SEQ ID NO: 492-SEQ ID NO: 759, SEQ ID NO: 761, SEQ ID NO: 763, SEQ ID NO: 765-SEQ ID NO: 818, SEQ ID NO: 820-SEQ ID NO: 846, SEQ ID NO: 848-SEQ ID NO: 896, SEQ ID NO: 898-SEQ ID NO: 963, SEQ ID NO: 966-SEQ ID NO: 982, SEQ ID NO: 1037, SEQ ID NO: 1038, SEQ ID NO: 1041-SEQ ID NO: 1087, SEQ ID NO: 1090 and SEQ ID NO: 1296-SEQ ID NO: 1298.

- 15           16.     A vaccine composition of claim 15, further comprising a pharmaceutically acceptable carrier.

17.     A vaccine composition of claim 16, wherein the pharmaceutically acceptable carrier is an adjuvant.

- 20           18.     A method of treating a subject for *H. pylori* infection comprising administering to a subject a vaccine composition of claim 15, such that treatment of *H. pylori* infection occurs.

- 25           19.     A method of claim 18, wherein the treatment is a prophylactic treatment.

20.     A method of claim 18, wherein the treatment is a therapeutic treatment.

21.     An isolated nucleic acid comprising a nucleotide sequence encoding an *H. pylori* cell envelope polypeptide or a fragment thereof, said nucleic acid selected from the group consisting of SEQ ID NO: 255, SEQ ID NO: 263, SEQ ID NO: 266, SEQ ID NO: 277, SEQ ID NO: 280, SEQ ID NO: 285, SEQ ID NO: 292, SEQ ID NO: 294, SEQ ID NO: 299, SEQ ID NO: 311, SEQ ID NO: 312, SEQ ID NO: 313, SEQ ID NO: 321, SEQ ID NO: 327, SEQ ID NO: 329, SEQ ID NO: 331, SEQ ID NO: 353, SEQ ID NO: 364, SEQ ID NO: 366, SEQ ID NO: 368, SEQ ID NO: 375, SEQ ID NO: 384, SEQ ID NO: 391, SEQ ID NO: 392, SEQ ID NO: 397, SEQ ID NO: 398, SEQ ID NO: 402, SEQ ID NO: 404, SEQ ID NO: 409, SEQ ID NO: 410, SEQ ID NO: 412, SEQ ID

- 1113 -

NO: 427, SEQ ID NO: 433, SEQ ID NO: 434, SEQ ID NO: 441, SEQ ID NO: 444, SEQ ID NO: 445, SEQ ID NO: 449, SEQ ID NO: 450, SEQ ID NO: 452, SEQ ID NO: 453, SEQ ID NO: 466, SEQ ID NO: 468, SEQ ID NO: 469, SEQ ID NO: 983, SEQ ID NO: 989, SEQ ID NO: 1008, SEQ ID NO: 1011, SEQ ID NO: 1014, SEQ ID NO: 1015, SEQ ID NO: 1029, SEQ ID NO: 1032, SEQ ID NO: 259, SEQ ID NO: 286, SEQ ID NO: 326, SEQ ID NO: 374, SEQ ID NO: 399, SEQ ID NO: 422, SEQ ID NO: 454, SEQ ID NO: 465, SEQ ID NO: 998, SEQ ID NO: 1009, SEQ ID NO: 1023, SEQ ID NO: 1294, SEQ ID NO: 1295, SEQ ID NO: 319, SEQ ID NO: 325, SEQ ID NO: 425, SEQ ID NO: 437, SEQ ID NO: 438, SEQ ID NO: 447, SEQ ID NO: 448, SEQ ID NO: 467, SEQ ID NO: 996, SEQ ID NO: 1027, SEQ ID NO: 1031, SEQ ID NO: 254, SEQ ID NO: 352, SEQ ID NO: 415, SEQ ID NO: 1019, SEQ ID NO: 381, SEQ ID NO: 389, SEQ ID NO: 1010, SEQ ID NO: 1012, SEQ ID NO: 354, SEQ ID NO: 372, SEQ ID NO: 400, SEQ ID NO: 421, SEQ ID NO: 1022, SEQ ID NO: 463, SEQ ID NO: 281, SEQ ID NO: 988, SEQ ID NO: 411, SEQ ID NO: 407, SEQ ID NO: 1017, SEQ ID NO: 290, SEQ ID NO: 417, SEQ ID NO: 430, SEQ ID NO: 992, SEQ ID NO: 1025, SEQ ID NO: 477, SEQ ID NO: 414, SEQ ID NO: 253, SEQ ID NO: 293, SEQ ID NO: 334, SEQ ID NO: 343, SEQ ID NO: 418, SEQ ID NO: 424, and SEQ ID NO: 443.

22. The purified nucleic acid of claim 21, wherein said *H. pylori* cell envelope polypeptide or a fragment thereof is an *H. pylori* outer membrane polypeptide or a fragment thereof encoded by the nucleic acid selected from the group consisting of SEQ ID NO: 255, SEQ ID NO: 263, SEQ ID NO: 266, SEQ ID NO: 277, SEQ ID NO: 280, SEQ ID NO: 285, SEQ ID NO: 292, SEQ ID NO: 294, SEQ ID NO: 299, SEQ ID NO: 311, SEQ ID NO: 312, SEQ ID NO: 313, SEQ ID NO: 321, SEQ ID NO: 327, SEQ ID NO: 329, SEQ ID NO: 331, SEQ ID NO: 353, SEQ ID NO: 364, SEQ ID NO: 366, SEQ ID NO: 368, SEQ ID NO: 375, SEQ ID NO: 384, SEQ ID NO: 391, SEQ ID NO: 392, SEQ ID NO: 397, SEQ ID NO: 398, SEQ ID NO: 402, SEQ ID NO: 404, SEQ ID NO: 409, SEQ ID NO: 410, SEQ ID NO: 412, SEQ ID NO: 427, SEQ ID NO: 433, SEQ ID NO: 434, SEQ ID NO: 441, SEQ ID NO: 444, SEQ ID NO: 445, SEQ ID NO: 449, SEQ ID NO: 450, SEQ ID NO: 452, SEQ ID NO: 453, SEQ ID NO: 466, SEQ ID NO: 468, SEQ ID NO: 469, SEQ ID NO: 983, SEQ ID NO: 989, SEQ ID NO: 1008, SEQ ID NO: 1011, SEQ ID NO: 1014, SEQ ID NO: 1015, SEQ ID NO: 1029, SEQ ID NO: 1032, SEQ ID NO: 259, SEQ ID NO: 286, SEQ ID NO: 326, SEQ ID NO: 374, SEQ ID NO: 399, SEQ ID NO: 422, SEQ ID NO: 454, SEQ ID NO: 465, SEQ ID NO: 998, SEQ ID NO: 1009, SEQ ID NO: 1023, SEQ ID NO: 1294, SEQ ID NO: 1295, SEQ ID NO: 319, SEQ ID NO: 325, SEQ ID NO: 425, SEQ ID NO: 437, SEQ ID NO: 438, SEQ ID

- 1114 -

NO: 447, SEQ ID NO: 448, SEQ ID NO: 467, SEQ ID NO: 996, SEQ ID NO: 1027, SEQ ID NO: 1031, SEQ ID NO: 254, SEQ ID NO: 352, SEQ ID NO: 415, SEQ ID NO: 1019, SEQ ID NO: 381, SEQ ID NO: 389, SEQ ID NO: 1010, and SEQ ID NO: 1012.

5           23.     he purified nucleic acid of claim 22, wherein said *H. pylori* outer  
membrane polypeptide or a fragment thereof is an *H. pylori* polypeptide having a  
terminal phenylalanine residue or a fragment thereof encoded by the nucleic acid  
selected from the group consisting of SEQ ID NO: 255, SEQ ID NO: 263, SEQ ID NO:  
266, SEQ ID NO: 277, SEQ ID NO: 280, SEQ ID NO: 285, SEQ ID NO: 292, SEQ ID  
10 NO: 294, SEQ ID NO: 299, SEQ ID NO: 311, SEQ ID NO: 312, SEQ ID NO: 313, SEQ  
ID NO: 321, SEQ ID NO: 327, SEQ ID NO: 329, SEQ ID NO: 331, SEQ ID NO: 353,  
SEQ ID NO: 364, SEQ ID NO: 366, SEQ ID NO: 368, SEQ ID NO: 375, SEQ ID NO:  
384, SEQ ID NO: 391, SEQ ID NO: 392, SEQ ID NO: 397, SEQ ID NO: 398, SEQ ID  
NO: 402, SEQ ID NO: 404, SEQ ID NO: 409, SEQ ID NO: 410, SEQ ID NO: 412, SEQ  
15 ID NO: 427, SEQ ID NO: 433, SEQ ID NO: 434, SEQ ID NO: 441, SEQ ID NO: 444,  
SEQ ID NO: 445, SEQ ID NO: 449, SEQ ID NO: 450, SEQ ID NO: 452, SEQ ID NO:  
453, SEQ ID NO: 466, SEQ ID NO: 468, SEQ ID NO: 469, SEQ ID NO: 983, SEQ ID  
NO: 989, SEQ ID NO: 1008, SEQ ID NO: 1011, SEQ ID NO: 1014, SEQ ID NO: 1015,  
SEQ ID NO: 1029, and SEQ ID NO: 1032.

20

24.     The purified nucleic acid of claim 22, wherein said *H. pylori* outer  
membrane polypeptide or a fragment thereof is an *H. pylori* polypeptide having a C-  
terminal tyrosine cluster or a fragment thereof encoded by the nucleic acid selected from  
the group consisting of SEQ ID NO: 286, SEQ ID NO: 326, SEQ ID NO: 374, SEQ ID  
25 NO: 399, SEQ ID NO: 422, SEQ ID NO: 454, SEQ ID NO: 465, SEQ ID NO: 998, SEQ  
ID NO: 1009, SEQ ID NO: 1023, SEQ ID NO: 1294, and SEQ ID NO: 1295.

25.     The purified nucleic acid of claim 22, wherein said *H. pylori* outer  
membrane polypeptide or a fragment thereof is an *H. pylori* polypeptide having a  
30 terminal phenylalanine residue and a C-terminal tyrosine cluster or a fragment thereof  
encoded by the nucleic acid selected from the group consisting of SEQ ID NO: 319,  
SEQ ID NO: 325, SEQ ID NO: 425, SEQ ID NO: 437, SEQ ID NO: 438, SEQ ID NO:  
447, SEQ ID NO: 448, SEQ ID NO: 467, SEQ ID NO: 996, SEQ ID NO: 1027, and  
SEQ ID NO: 1031.

35

- 1115 -

26. The purified nucleic acid of claim 21, wherein said *H. pylori* cell envelope polypeptide or a fragment thereof is an *H. pylori* inner membrane polypeptide or a fragment thereof encoded by the nucleic acid selected from the group consisting of SEQ ID NO: 354, SEQ ID NO: 372, SEQ ID NO: 400, SEQ ID NO: 421, SEQ ID NO: 1022, SEQ ID NO: 463, SEQ ID NO: 281, SEQ ID NO: 988, SEQ ID NO: 411, SEQ ID NO: 407, SEQ ID NO: 1017, SEQ ID NO: 290, SEQ ID NO: 417, SEQ ID NO: 430, SEQ ID NO: 992, and SEQ ID NO: 1025.

27. The purified nucleic acid of claim 26, wherein said *H. pylori* inner membrane polypeptide or a fragment thereof is an *H. pylori* polypeptide involved in outer membrane and cell wall synthesis or a fragment thereof encoded by the nucleic acid comprising a nucleotide sequence of SEQ ID NO: 354.

28. The purified nucleic acid of claim 26, wherein said *H. pylori* inner membrane polypeptide or a fragment thereof is an *H. pylori* polypeptide involved in energy conversion or a fragment thereof encoded by the nucleic acid selected from the group consisting of SEQ ID NO: 372, SEQ ID NO: 400, SEQ ID NO: 421, and SEQ ID NO: 1022.

29. The purified nucleic acid of claim 26, wherein said *H. pylori* inner membrane polypeptide or a fragment thereof is an *H. pylori* polypeptide involved in cofactor metabolism or a fragment thereof encoded by the nucleic acid comprising a nucleotide sequence of SEQ ID NO: 463.

30. The purified nucleic acid of claim 26, wherein said *H. pylori* inner membrane polypeptide or a fragment thereof is an *H. pylori* polypeptide involved in secretion and adhesion or a fragment thereof encoded by the nucleic acid selected from the group consisting of SEQ ID NO: 281 and SEQ ID NO: 988.

31. The purified nucleic acid of claim 26, wherein said *H. pylori* inner membrane polypeptide or a fragment thereof is an *H. pylori* polypeptide involved in transport or a fragment thereof encoded by the nucleic acid selected from the group consisting of SEQ ID NO: 407 and SEQ ID NO: 1017.

32. The purified nucleic acid of claim 21, wherein said *H. pylori* cell envelope polypeptide or a fragment thereof is an *H. pylori* flagellar polypeptide or a

- 1116 -

fragment thereof encoded by the nucleic acid comprising a nucleotide sequence of SEQ ID NO: 477.

33. The purified nucleic acid of claim 21, wherein said *H. pylori* cell  
5 envelope polypeptide or a fragment thereof is an *H. pylori* transport polypeptide or a fragment thereof encoded by the nucleic acid comprising a nucleotide sequence of SEQ ID NO: 414.

34. A recombinant expression vector comprising the nucleic acid of claim 21  
10 operably linked to a transcription regulatory element.

35. A cell comprising a recombinant expression vector of claim 34.

36. A method for producing an *H. pylori* polypeptide comprising culturing a  
15 cell of claim 35 under conditions that permit expression of the polypeptide.

37. An isolated nucleic acid comprising a nucleotide sequence encoding an  
*H. pylori* secreted polypeptide or a fragment thereof, said nucleic acid selected from the  
group consisting of SEQ ID NO: 355, SEQ ID NO: 1006, SEQ ID NO: 257, SEQ ID  
20 NO: 258, SEQ ID NO: 260, SEQ ID NO: 261, SEQ ID NO: 264, SEQ ID NO: 265, SEQ  
ID NO: 268, SEQ ID NO: 270, SEQ ID NO: 272, SEQ ID NO: 274, SEQ ID NO: 275,  
SEQ ID NO: 276, SEQ ID NO: 279, SEQ ID NO: 283, SEQ ID NO: 284, SEQ ID NO:  
287, SEQ ID NO: 288, SEQ ID NO: 289, SEQ ID NO: 291, SEQ ID NO: 295, SEQ ID  
NO: 296, SEQ ID NO: 297, SEQ ID NO: 298, SEQ ID NO: 300, SEQ ID NO: 301, SEQ  
25 ID NO: 302, SEQ ID NO: 303, SEQ ID NO: 304, SEQ ID NO: 305, SEQ ID NO: 314,  
SEQ ID NO: 315, SEQ ID NO: 323, SEQ ID NO: 338, SEQ ID NO: 342, SEQ ID NO:  
348, SEQ ID NO: 349, SEQ ID NO: 356, SEQ ID NO: 358, SEQ ID NO: 359, SEQ ID  
NO: 360, SEQ ID NO: 361, SEQ ID NO: 362, SEQ ID NO: 363, SEQ ID NO: 367, SEQ  
ID NO: 370, SEQ ID NO: 371, SEQ ID NO: 373, SEQ ID NO: 377, SEQ ID NO: 378,  
30 SEQ ID NO: 379, SEQ ID NO: 380, SEQ ID NO: 388, SEQ ID NO: 390, SEQ ID NO:  
394, SEQ ID NO: 395, SEQ ID NO: 396, SEQ ID NO: 401, SEQ ID NO: 403, SEQ ID  
NO: 405, SEQ ID NO: 408, SEQ ID NO: 420, SEQ ID NO: 426, SEQ ID NO: 428, SEQ  
ID NO: 429, SEQ ID NO: 432, SEQ ID NO: 439, SEQ ID NO: 442, SEQ ID NO: 451,  
SEQ ID NO: 471, SEQ ID NO: 478, SEQ ID NO: 488, SEQ ID NO: 987, SEQ ID NO:  
35 990, SEQ ID NO: 991, SEQ ID NO: 993, SEQ ID NO: 1001, SEQ ID NO: 1002, SEQ

- 1117 -

ID NO: 1007, SEQ ID NO: 1013, SEQ ID NO: 1016, SEQ ID NO: 1018, SEQ ID NO: 1021, and SEQ ID NO: 1026.

38. The purified nucleic acid of claim 37, wherein said *H. pylori* secreted polypeptide or a fragment thereof is an *H. pylori* polypeptide involved in secretion and adhesion or a fragment thereof encoded by the nucleic acid selected from the group consisting of SEQ ID NO: 355 and SEQ ID NO: 1006.

39. A recombinant expression vector comprising the nucleic acid of claim 38 operably linked to a transcription regulatory element.

40. A cell comprising a recombinant expression vector of claim 39.

41. A method for producing an *H. pylori* polypeptide comprising culturing a cell of claim 40 under conditions that permit expression of the polypeptide.

42. An isolated nucleic acid comprising a nucleotide sequence encoding an *H. pylori* cytoplasmic polypeptide or a fragment thereof, said nucleic acid selected from the group consisting of SEQ ID NO: 470, SEQ ID NO: 1033, SEQ ID NO: 357, SEQ ID NO: 457, SEQ ID NO: 461, SEQ ID NO: 1030, SEQ ID NO: 345, SEQ ID NO: 383, SEQ ID NO: 387, SEQ ID NO: 455, SEQ ID NO: 1003, SEQ ID NO: 351, SEQ ID NO: 416, SEQ ID NO: 278, SEQ ID NO: 335, SEQ ID NO: 346, SEQ ID NO: 350, SEQ ID NO: 419, SEQ ID NO: 460, SEQ ID NO: 472, SEQ ID NO: 1000, SEQ ID NO: 1004, SEQ ID NO: 1020, SEQ ID NO: 1293, SEQ ID NO: 318, SEQ ID NO: 322, SEQ ID NO: 324, SEQ ID NO: 330, SEQ ID NO: 347, SEQ ID NO: 440, SEQ ID NO: 446, SEQ ID NO: 464, SEQ ID NO: 490, SEQ ID NO: 491, SEQ ID NO: 995, SEQ ID NO: 997, SEQ ID NO: 1005, SEQ ID NO: 1028.

43. The purified nucleic acid of claim 42, wherein said *H. pylori* cytoplasmic polypeptide or a fragment thereof is an *H. pylori* polypeptide involved in energy conversion or a fragment thereof encoded by the nucleic acid selected from the group consisting of SEQ ID NO: 470 and SEQ ID NO: 1033.

44. The purified nucleic acid of claim 42, wherein said *H. pylori* cytoplasmic polypeptide or a fragment thereof is an *H. pylori* polypeptide involved in amino acid

- 1118 -

metabolism and transport or a fragment thereof encoded by the nucleic acid selected from the group consisting of SEQ ID NO: 357 and SEQ ID NO: 457.

45. The purified nucleic acid of claim 42, wherein said *H. pylori* cytoplasmic polypeptide or a fragment thereof is an *H. pylori* polypeptide involved in nucleotide metabolism and transport or a fragment thereof encoded by the nucleic acid selected from the group consisting of SEQ ID NO: 461 and SEQ ID NO: 1030.

46. The purified nucleic acid of claim 42, wherein said *H. pylori* cytoplasmic polypeptide or a fragment thereof is an *H. pylori* polypeptide involved in cofactor metabolism or a fragment thereof encoded by the nucleic acid selected from the group consisting of SEQ ID NO: 345, SEQ ID NO: 383, SEQ ID NO: 387, SEQ ID NO: 455, and SEQ ID NO: 1003.

47. The purified nucleic acid of claim 42, wherein said *H. pylori* cytoplasmic polypeptide or a fragment thereof is an *H. pylori* polypeptide involved in lipid metabolism or a fragment thereof encoded by the nucleic acid selected from the group consisting of SEQ ID NO: 351 and SEQ ID NO: 416.

48. The purified nucleic acid of claim 42, wherein said *H. pylori* cytoplasmic polypeptide or a fragment thereof is an *H. pylori* polypeptide involved in genome replication, transcription, recombination and repair or a fragment thereof encoded by the nucleic acid selected from the group consisting of SEQ ID NO: 278, SEQ ID NO: 335, SEQ ID NO: 346, SEQ ID NO: 350, SEQ ID NO: 419, SEQ ID NO: 460, SEQ ID NO: 472, SEQ ID NO: 1000, SEQ ID NO: 1004, SEQ ID NO: 1020, and SEQ ID NO: 1293.

49. A recombinant expression vector comprising the nucleic acid of claim 42 operably linked to a transcription regulatory element.

50. A cell comprising a recombinant expression vector of claim 49.

51. A method for producing an *H. pylori* polypeptide comprising culturing a cell of claim 50 under conditions that permit expression of the polypeptide.

52. An isolated nucleic acid comprising a nucleotide sequence encoding an *H. pylori* cellular polypeptide or a fragment thereof, said nucleic acid selected from the

- 1119 -

group consisting of SEQ ID NO: 256, SEQ ID NO: 267, SEQ ID NO: 282, SEQ ID NO: 306, SEQ ID NO: 307, SEQ ID NO: 308, SEQ ID NO: 309, SEQ ID NO: 310, SEQ ID NO: 316, SEQ ID NO: 317, SEQ ID NO: 332, SEQ ID NO: 333, SEQ ID NO: 336, SEQ ID NO: 337, SEQ ID NO: 339, SEQ ID NO: 340, SEQ ID NO: 341, SEQ ID NO: 344,  
5 SEQ ID NO: 369, SEQ ID NO: 376, SEQ ID NO: 382, SEQ ID NO: 386, SEQ ID NO: 423, SEQ ID NO: 431, SEQ ID NO: 435, SEQ ID NO: 436, SEQ ID NO: 458, SEQ ID NO: 462, SEQ ID NO: 475, SEQ ID NO: 476, SEQ ID NO: 479, SEQ ID NO: 480, SEQ ID NO: 481, SEQ ID NO: 482, SEQ ID NO: 483, SEQ ID NO: 484, SEQ ID NO: 485, SEQ ID NO: 486, SEQ ID NO: 487, SEQ ID NO: 489, SEQ ID NO: 984, SEQ ID NO: 994, SEQ ID NO: 1024, and SEQ ID NO: 1036.

53. A recombinant expression vector comprising the nucleic acid of claim 52 operably linked to a transcription regulatory element.

15 54. A cell comprising a recombinant expression vector of claim 53.

55. A method for producing an *H. pylori* polypeptide comprising culturing a cell of claim 54 under conditions that permit expression of the polypeptide.

20 56. A vaccine composition for prevention or treatment of an *H. pylori* infection comprising an effective amount of a nucleic acid of claim 21.

57. A vaccine composition of claim 56, further comprising a pharmaceutically acceptable carrier.

25

58. A vaccine composition of claim 57, wherein the pharmaceutically acceptable carrier is an adjuvant.

59. A method of treating a subject for *H. pylori* infection comprising  
30 administering to a subject a vaccine composition of claim 56, such that treatment of *H. pylori* infection occurs.

60. A method of claim 59, wherein the treatment is a prophylactic treatment.

35 61. A method of claim 59, wherein the treatment is a therapeutic treatment.

- 1120 -

62. A vaccine composition for prevention or treatment of an *H. pylori* infection comprising an effective amount of a nucleic acid of claim 37.

5 63. A vaccine composition of claim 62, further comprising a pharmaceutically acceptable carrier.

64. A vaccine composition of claim 63, wherein the pharmaceutically acceptable carrier is an adjuvant.

10 65. A method of treating a subject for *H. pylori* infection comprising administering to a subject a vaccine composition of claim 62, such that treatment of *H. pylori* infection occurs.

15 66. A method of claim 65, wherein the treatment is a prophylactic treatment.

67. A method of claim 65, wherein the treatment is a therapeutic treatment.

20 68. A vaccine composition for prevention or treatment of an *H. pylori* infection comprising an effective amount of a nucleic acid of claim 42.

69. A vaccine composition of claim 68, further comprising a pharmaceutically acceptable carrier.

25 70. A vaccine composition of claim 69, wherein the pharmaceutically acceptable carrier is an adjuvant.

30 71. A method of treating a subject for *H. pylori* infection comprising administering to a subject a vaccine composition of claim 68, such that treatment of *H. pylori* infection occurs.

72. A method of claim 71, wherein the treatment is a prophylactic treatment.

73. A method of claim 71, wherein the treatment is a therapeutic treatment.

35 74. A vaccine composition for prevention or treatment of an *H. pylori* infection comprising an effective amount of a nucleic acid of claim 52.

- 1121 -

75. A vaccine composition of claim 74, further comprising a pharmaceutically acceptable carrier.

5 76. A vaccine composition of claim 75, wherein the pharmaceutically acceptable carrier is an adjuvant.

77. A method of treating a subject for *H. pylori* infection comprising administering to a subject a vaccine composition of claim 74, such that treatment of *H.*  
10 *pylori* infection occurs.

78. A method of claim 77, wherein the treatment is a prophylactic treatment.

79. A method of claim 77, wherein the treatment is a therapeutic treatment.  
15

80. A purified *H. pylori* cell envelope polypeptide or a fragment thereof, wherein said polypeptide is selected from the group consisting of SEQ ID NO: 746, SEQ ID NO: 754, SEQ ID NO: 757, SEQ ID NO: 768, SEQ ID NO: 771, SEQ ID NO: 776, SEQ ID NO: 783, SEQ ID NO: 785, SEQ ID NO: 790, SEQ ID NO: 802, SEQ ID  
20 NO: 803, SEQ ID NO: 804, SEQ ID NO: 812, SEQ ID NO: 818, SEQ ID NO: 820, SEQ ID NO: 882, SEQ ID NO: 844, SEQ ID NO: 855, SEQ ID NO: 857, SEQ ID NO: 859, SEQ ID NO: 866, SEQ ID NO: 875, SEQ ID NO: 882, SEQ ID NO: 883, SEQ ID NO: 888, SEQ ID NO: 889, SEQ ID NO: 893, SEQ ID NO: 895, SEQ ID NO: 900, SEQ ID NO: 901, SEQ ID NO: 903, SEQ ID NO: 918, SEQ ID NO: 924, SEQ ID NO: 925, SEQ  
25 ID NO: 932, SEQ ID NO: 935, SEQ ID NO: 936, SEQ ID NO: 940, SEQ ID NO: 941, SEQ ID NO: 943, SEQ ID NO: 944, SEQ ID NO: 957, SEQ ID NO: 959, SEQ ID NO: 960, SEQ ID NO: 1037, SEQ ID NO: 1043, SEQ ID NO: 1062, SEQ ID NO: 1065, SEQ ID NO: 1068, SEQ ID NO: 1069, SEQ ID NO: 1083, SEQ ID NO: 1086, SEQ ID NO: 750, SEQ ID NO: 777, SEQ ID NO: 817, SEQ ID NO: 865, SEQ ID NO: 890, SEQ ID  
30 NO: 913, SEQ ID NO: 945, SEQ ID NO: 956, SEQ ID NO: 1052, SEQ ID NO: 1063, SEQ ID NO: 1077, SEQ ID NO: 1297, SEQ ID NO: 1298, SEQ ID NO: 810, SEQ ID NO: 816, SEQ ID NO: 916, SEQ ID NO: 928, SEQ ID NO: 929, SEQ ID NO: 938, SEQ ID NO: 939, SEQ ID NO: 958, SEQ ID NO: 1050, SEQ ID NO: 1081, SEQ ID NO: 1085, SEQ ID NO: 745, SEQ ID NO: 843, SEQ ID NO: 906, SEQ ID NO: 1073, SEQ  
35 ID NO: 872, SEQ ID NO: 880, SEQ ID NO: 1064, SEQ ID NO: 1066, SEQ ID NO: 845, SEQ ID NO: 863, SEQ ID NO: 891, SEQ ID NO: 912, SEQ ID NO: 1076, SEQ ID

- 1122 -

NO: 954, SEQ ID NO: 772, SEQ ID NO: 1042, SEQ ID NO: 902, SEQ ID NO: 898,  
 SEQ ID NO: 1071, SEQ ID NO: 781, SEQ ID NO: 908, SEQ ID NO: 921, SEQ ID NO:  
 1046, SEQ ID NO: 1079, SEQ ID NO: 968, SEQ ID NO: 905, SEQ ID NO: 744, SEQ  
 ID NO: 784, SEQ ID NO: 825, SEQ ID NO: 834, SEQ ID NO: 909, SEQ ID NO: 915,  
 5 and SEQ ID NO: 934.

81. The purified polypeptide of claim 80, wherein said *H. pylori* cell  
 envelope polypeptide or a fragment thereof is an *H. pylori* outer membrane polypeptide  
 or a fragment thereof selected from the group consisting of SEQ ID NO: 746, SEQ ID  
 10 NO: 754, SEQ ID NO: 757, SEQ ID NO: 768, SEQ ID NO: 771, SEQ ID NO: 776, SEQ  
 ID NO: 783, SEQ ID NO: 785, SEQ ID NO: 790, SEQ ID NO: 802, SEQ ID NO: 803,  
 SEQ ID NO: 804, SEQ ID NO: 812, SEQ ID NO: 818, SEQ ID NO: 820, SEQ ID NO:  
 882, SEQ ID NO: 844, SEQ ID NO: 855, SEQ ID NO: 857, SEQ ID NO: 859, SEQ ID  
 NO: 866, SEQ ID NO: 875, SEQ ID NO: 882, SEQ ID NO: 883, SEQ ID NO: 888, SEQ  
 15 ID NO: 889, SEQ ID NO: 893, SEQ ID NO: 895, SEQ ID NO: 900, SEQ ID NO: 901,  
 SEQ ID NO: 903, SEQ ID NO: 918, SEQ ID NO: 924, SEQ ID NO: 925, SEQ ID NO:  
 932, SEQ ID NO: 935, SEQ ID NO: 936, SEQ ID NO: 940, SEQ ID NO: 941, SEQ ID  
 NO: 943, SEQ ID NO: 944, SEQ ID NO: 957, SEQ ID NO: 959, SEQ ID NO: 960, SEQ  
 ID NO: 1037, SEQ ID NO: 1043, SEQ ID NO: 1062, SEQ ID NO: 1065, SEQ ID NO:  
 20 1068, SEQ ID NO: 1069, SEQ ID NO: 1083, SEQ ID NO: 1086, SEQ ID NO: 750, SEQ  
 ID NO: 777, SEQ ID NO: 817, SEQ ID NO: 865, SEQ ID NO: 890, SEQ ID NO: 913,  
 SEQ ID NO: 945, SEQ ID NO: 956, SEQ ID NO: 1052, SEQ ID NO: 1063, SEQ ID  
 NO: 1077, SEQ ID NO: 1297, SEQ ID NO: 1298, SEQ ID NO: 810, SEQ ID NO: 816,  
 SEQ ID NO: 916, SEQ ID NO: 928, SEQ ID NO: 929, SEQ ID NO: 938, SEQ ID NO:  
 25 939, SEQ ID NO: 958, SEQ ID NO: 1050, SEQ ID NO: 1081, SEQ ID NO: 1085, SEQ  
 ID NO: 745, SEQ ID NO: 843, SEQ ID NO: 906, SEQ ID NO: 1073, SEQ ID NO: 872,  
 SEQ ID NO: 880, SEQ ID NO: 1064, and SEQ ID NO: 1066.

82. The purified polypeptide of claim 81, wherein said *H. pylori* outer  
 30 membrane polypeptide or a fragment thereof is an *H. pylori* polypeptide having a  
 terminal phenylalanine residue or a fragment thereof selected from the group consisting  
 of SEQ ID NO: 746, SEQ ID NO: 754, SEQ ID NO: 757, SEQ ID NO: 768, SEQ ID  
 NO: 771, SEQ ID NO: 776, SEQ ID NO: 783, SEQ ID NO: 785, SEQ ID NO: 790, SEQ  
 ID NO: 802, SEQ ID NO: 803, SEQ ID NO: 804, SEQ ID NO: 812, SEQ ID NO: 818,  
 35 SEQ ID NO: 820, SEQ ID NO: 882, SEQ ID NO: 844, SEQ ID NO: 855, SEQ ID NO:  
 857, SEQ ID NO: 859, SEQ ID NO: 866, SEQ ID NO: 875, SEQ ID NO: 882, SEQ ID

- 1123 -

NO: 883, SEQ ID NO: 888, SEQ ID NO: 889, SEQ ID NO: 893, SEQ ID NO: 895, SEQ ID NO: 900, SEQ ID NO: 901, SEQ ID NO: 903, SEQ ID NO: 918, SEQ ID NO: 924, SEQ ID NO: 925, SEQ ID NO: 932, SEQ ID NO: 935, SEQ ID NO: 936, SEQ ID NO: 940, SEQ ID NO: 941, SEQ ID NO: 943, SEQ ID NO: 944, SEQ ID NO: 957, SEQ ID NO: 959, SEQ ID NO: 960, SEQ ID NO: 1037, SEQ ID NO: 1043, SEQ ID NO: 1062, SEQ ID NO: 1065, SEQ ID NO: 1068, SEQ ID NO: 1069, SEQ ID NO: 1083, and SEQ ID NO: 1086.

83. The purified polypeptide of claim 81, wherein said *H. pylori* outer membrane polypeptide or a fragment thereof is an *H. pylori* polypeptide a C-terminal tyrosine cluster or a fragment thereof selected from the group consisting of SEQ ID NO: 746, SEQ ID NO: 754, SEQ ID NO: 757, SEQ ID NO: 768, SEQ ID NO: 771, SEQ ID NO: 776, SEQ ID NO: 783, SEQ ID NO: 785, SEQ ID NO: 790, SEQ ID NO: 802, SEQ ID NO: 803, SEQ ID NO: 804, SEQ ID NO: 812, SEQ ID NO: 818, SEQ ID NO: 820, SEQ ID NO: 882, SEQ ID NO: 844, SEQ ID NO: 855, SEQ ID NO: 857, SEQ ID NO: 859, SEQ ID NO: 866, SEQ ID NO: 875, SEQ ID NO: 882, SEQ ID NO: 883, SEQ ID NO: 888, SEQ ID NO: 889, SEQ ID NO: 893, SEQ ID NO: 895, SEQ ID NO: 900, SEQ ID NO: 901, SEQ ID NO: 903, SEQ ID NO: 918, SEQ ID NO: 924, SEQ ID NO: 925, SEQ ID NO: 932, SEQ ID NO: 935, SEQ ID NO: 936, SEQ ID NO: 940, SEQ ID NO: 941, SEQ ID NO: 943, SEQ ID NO: 944, SEQ ID NO: 957, SEQ ID NO: 959, SEQ ID NO: 960, SEQ ID NO: 1037, SEQ ID NO: 1043, SEQ ID NO: 1062, SEQ ID NO: 1065, SEQ ID NO: 1068, SEQ ID NO: 1069, SEQ ID NO: 1083, and SEQ ID NO: 1086.

84. The purified polypeptide of claim 81, wherein said *H. pylori* outer membrane polypeptide or a fragment thereof is an *H. pylori* polypeptide having a terminal phenylalanine residue and a C-terminal tyrosine cluster or a fragment thereof selected from the group consisting of SEQ ID NO: 810, SEQ ID NO: 816, SEQ ID NO: 916, SEQ ID NO: 928, SEQ ID NO: 929, SEQ ID NO: 938, SEQ ID NO: 939, SEQ ID NO: 958, SEQ ID NO: 1050, SEQ ID NO: 1081, and SEQ ID NO: 1085.

85. The purified polypeptide of claim 80, wherein said *H. pylori* cell envelope polypeptide or a fragment thereof is an *H. pylori* inner membrane polypeptide or a fragment thereof selected from the group consisting of SEQ ID NO: 845, SEQ ID NO: 863, SEQ ID NO: 891, SEQ ID NO: 912, SEQ ID NO: 1076, SEQ ID NO: 954, SEQ ID NO: 772, SEQ ID NO: 1042, SEQ ID NO: 902, SEQ ID NO: 898, SEQ ID NO:

- 1124 -

1071, SEQ ID NO: 781, SEQ ID NO: 908, SEQ ID NO: 921, SEQ ID NO: 1046, SEQ ID NO: 1079.

86. The purified polypeptide of claim 85, wherein said *H. pylori* inner  
5 membrane polypeptide or a fragment thereof is an *H. pylori* polypeptide involved in outer membrane and cell wall synthesis or a fragment thereof comprising an amino acid sequence of SEQ ID NO: 845.

87. The purified polypeptide of claim 85, wherein said *H. pylori* inner  
10 membrane polypeptide or a fragment thereof is an *H. pylori* polypeptide involved in energy conversion or a fragment thereof selected from the group consisting of SEQ ID NO: 863, SEQ ID NO: 891, SEQ ID NO: 912, and SEQ ID NO: 1076.

88. The purified polypeptide of claim 85, wherein said *H. pylori* inner  
15 membrane polypeptide or a fragment thereof is an *H. pylori* polypeptide involved in cofactor metabolism or a fragment thereof comprising an amino acid sequence of SEQ ID NO: 954.

89. The purified polypeptide of claim 85, wherein said *H. pylori* inner  
20 membrane polypeptide or a fragment thereof is an *H. pylori* polypeptide involved in secretion and adhesion or a fragment thereof selected from the group consisting of SEQ ID NO: 772 and SEQ ID NO: 1042.

90. The purified polypeptide of claim 85, wherein said *H. pylori* inner  
25 membrane polypeptide or a fragment thereof is an *H. pylori* polypeptide involved in transport or a fragment thereof selected from the group consisting of SEQ ID NO: 898 and SEQ ID NO: 1071.

91. The purified polypeptide of claim 80, wherein said *H. pylori* cell  
30 envelope polypeptide or a fragment thereof is an *H. pylori* flagellar polypeptide or a fragment thereof comprising an amino acid sequence of SEQ ID NO: 968.

92. The purified polypeptide of claim 80, wherein said *H. pylori* cell  
35 envelope polypeptide or a fragment thereof is an *H. pylori* transport polypeptide or a fragment thereof comprising an amino acid sequence of SEQ ID NO: 905.

- 1125 -

93. A purified *H. pylori* cellular polypeptide or a fragment thereof, wherein said polypeptide is selected from the group consisting of SEQ ID NO: 747, SEQ ID NO: 758, SEQ ID NO: 773, SEQ ID NO: 797, SEQ ID NO: 798, SEQ ID NO: 799, SEQ ID NO: 800, SEQ ID NO: 801, SEQ ID NO: 807, SEQ ID NO: 808, SEQ ID NO: 823, SEQ ID NO: 824, SEQ ID NO: 827, SEQ ID NO: 828, SEQ ID NO: 830, SEQ ID NO: 831, SEQ ID NO: 832, SEQ ID NO: 835, SEQ ID NO: 860, SEQ ID NO: 867, SEQ ID NO: 873, SEQ ID NO: 877, SEQ ID NO: 914, SEQ ID NO: 922, SEQ ID NO: 926, SEQ ID NO: 927, SEQ ID NO: 949, SEQ ID NO: 953, SEQ ID NO: 966, SEQ ID NO: 967, SEQ ID NO: 970, SEQ ID NO: 971, SEQ ID NO: 972, SEQ ID NO: 973, SEQ ID NO: 974, SEQ ID NO: 975, SEQ ID NO: 976, SEQ ID NO: 977, SEQ ID NO: 978, SEQ ID NO: 980, SEQ ID NO: 1038, SEQ ID NO: 1048, SEQ ID NO: 1078, and SEQ ID NO: 1090.

94. A purified *H. pylori* secreted polypeptide or a fragment thereof, wherein said polypeptide is selected from the group consisting of SEQ ID NO: 846, SEQ ID NO: 1060, SEQ ID NO: 748, SEQ ID NO: 749, SEQ ID NO: 751, SEQ ID NO: 752, SEQ ID NO: 755, SEQ ID NO: 756, SEQ ID NO: 759, SEQ ID NO: 761, SEQ ID NO: 763, SEQ ID NO: 765, SEQ ID NO: 766, SEQ ID NO: 767, SEQ ID NO: 770, SEQ ID NO: 774, SEQ ID NO: 775, SEQ ID NO: 778, SEQ ID NO: 779, SEQ ID NO: 780, SEQ ID NO: 782, SEQ ID NO: 786, SEQ ID NO: 787, SEQ ID NO: 788, SEQ ID NO: 789, SEQ ID NO: 791, SEQ ID NO: 792, SEQ ID NO: 793, SEQ ID NO: 794, SEQ ID NO: 795, SEQ ID NO: 796, SEQ ID NO: 805, SEQ ID NO: 806, SEQ ID NO: 814, SEQ ID NO: 829, SEQ ID NO: 833, SEQ ID NO: 839, SEQ ID NO: 840, SEQ ID NO: 849, SEQ ID NO: 850, SEQ ID NO: 851, SEQ ID NO: 852, SEQ ID NO: 853, SEQ ID NO: 854, SEQ ID NO: 858, SEQ ID NO: 861, SEQ ID NO: 862, SEQ ID NO: 864, SEQ ID NO: 868, SEQ ID NO: 869, SEQ ID NO: 870, SEQ ID NO: 871, SEQ ID NO: 879, SEQ ID NO: 881, SEQ ID NO: 885, SEQ ID NO: 886, SEQ ID NO: 887, SEQ ID NO: 892, SEQ ID NO: 894, SEQ ID NO: 896, SEQ ID NO: 899, SEQ ID NO: 911, SEQ ID NO: 917, SEQ ID NO: 919, SEQ ID NO: 920, SEQ ID NO: 923, SEQ ID NO: 930, SEQ ID NO: 933, SEQ ID NO: 942, SEQ ID NO: 962, SEQ ID NO: 969, SEQ ID NO: 979, SEQ ID NO: 1041, SEQ ID NO: 1044, SEQ ID NO: 1045, SEQ ID NO: 1047, SEQ ID NO: 1055, SEQ ID NO: 1056, SEQ ID NO: 1061, SEQ ID NO: 1067, SEQ ID NO: 1070, SEQ ID NO: 1072, SEQ ID NO: 1075, and SEQ ID NO: 1080.

95. The purified polypeptide of claim 94, wherein said *H. pylori* secreted polypeptide or a fragment thereof is an *H. pylori* polypeptide involved in secretion and

- 1126 -

adhesion or a fragment thereof selected from the group consisting of SEQ ID NO: 846 and SEQ ID NO: 1060.

96. A purified *H. pylori* cytoplasmic polypeptide or a fragment thereof,  
5 wherein said polypeptide is selected from the group consisting of SEQ ID NO: 961, SEQ ID NO: 1087, SEQ ID NO: 848, SEQ ID NO: 948, SEQ ID NO: 952, SEQ ID NO: 1084, SEQ ID NO: 836, SEQ ID NO: 874, SEQ ID NO: 878, SEQ ID NO: 946, SEQ ID NO: 1057, SEQ ID NO: 842, SEQ ID NO: 907, SEQ ID NO: 769, SEQ ID NO: 826, SEQ ID NO: 837, SEQ ID NO: 841, SEQ ID NO: 910, SEQ ID NO: 951, SEQ ID NO:  
10 963, SEQ ID NO: 1054, SEQ ID NO: 1058, SEQ ID NO: 1074, SEQ ID NO: 1296, SEQ ID NO: 809, SEQ ID NO: 813, SEQ ID NO: 815, SEQ ID NO: 821, SEQ ID NO: 838, SEQ ID NO: 931, SEQ ID NO: 937, SEQ ID NO: 955, SEQ ID NO: 981, SEQ ID NO: 982, SEQ ID NO: 1049, SEQ ID NO: 1051, SEQ ID NO: 1059, and SEQ ID NO: 1082.

15 97. The purified polypeptide of claim 96, wherein said *H. pylori* cytoplasmic polypeptide or a fragment thereof is an *H. pylori* polypeptide involved in energy conversion or a fragment thereof selected from the group consisting of SEQ ID NO: 961 and SEQ ID NO: 1087.

20 98. The purified polypeptide of claim 96, wherein said *H. pylori* cytoplasmic polypeptide or a fragment thereof is an *H. pylori* polypeptide involved in amino acid metabolism and transport or a fragment thereof selected from the group consisting of SEQ ID NO: 848 and SEQ ID NO: 948.

25 99. The purified polypeptide of claim 96, wherein said *H. pylori* cytoplasmic polypeptide or a fragment thereof is an *H. pylori* polypeptide involved in nucleotide metabolism and transport or a fragment thereof selected from the group consisting of SEQ ID NO: 952 and SEQ ID NO: 1084.

30 100. The purified polypeptide of claim 96, wherein said *H. pylori* cytoplasmic polypeptide or a fragment thereof is an *H. pylori* polypeptide involved in cofactor metabolism or a fragment thereof selected from the group consisting of SEQ ID NO: 836, SEQ ID NO: 874, SEQ ID NO: 878, SEQ ID NO: 946, and SEQ ID NO: 1057.

35 101. The purified polypeptide of claim 96, wherein said *H. pylori* cytoplasmic polypeptide or a fragment thereof is an *H. pylori* polypeptide involved in lipid

- 1127 -

metabolism or a fragment thereof selected from the group consisting of SEQ ID NO: 842 and SEQ ID NO: 907.

102. The purified polypeptide of claim 96, wherein said *H. pylori* cytoplasmic  
5 polypeptide or a fragment thereof is an *H. pylori* polypeptide involved in genome replication, transcription, recombination and repair or a fragment thereof selected from the group consisting of SEQ ID NO: 769, SEQ ID NO: 826, SEQ ID NO: 837, SEQ ID NO: 841, SEQ ID NO: 910, SEQ ID NO: 951, SEQ ID NO: 963, SEQ ID NO: 1054, SEQ ID NO: 1058, SEQ ID NO: 1074, and SEQ ID NO: 1296.

10

103. A vaccine composition for prevention or treatment of an *H. pylori* infection comprising an effective amount of an *H. pylori* polypeptide of claim 80.

104. A vaccine composition of claim 103, further comprising a  
15 pharmaceutically acceptable carrier.

105. A vaccine composition of claim 104, wherein the pharmaceutically acceptable carrier is an adjuvant.

20 106. A method of treating a subject for *H. pylori* infection comprising administering to a subject a vaccine composition of claim 103, such that treatment of *H. pylori* infection occurs.

107. A method of claim 106, wherein the treatment is a prophylactic  
25 treatment.

108. A method of claim 106, wherein the treatment is a therapeutic treatment.

109. A vaccine composition for prevention or treatment of an *H. pylori*  
30 infection comprising an effective amount of an *H. pylori* polypeptide of claim 93.

110. A vaccine composition of claim 109, further comprising a pharmaceutically acceptable carrier.

35 111. A vaccine composition of claim 110, wherein the pharmaceutically acceptable carrier is an adjuvant.

- 1128 -

112. A method of treating a subject for *H. pylori* infection comprising administering to a subject a vaccine composition of claim 109, such that treatment of *H. pylori* infection occurs.

5

113. A method of claim 112, wherein the treatment is a prophylactic treatment.

114. A method of claim 112, wherein the treatment is a therapeutic treatment.

10

115. A vaccine composition for prevention or treatment of an *H. pylori* infection comprising an effective amount of an *H. pylori* polypeptide of claim 94.

116. A vaccine composition of claim 115, further comprising a  
15 pharmaceutically acceptable carrier.

117. A vaccine composition of claim 116, wherein the pharmaceutically acceptable carrier is an adjuvant.

20 118. A method of treating a subject for *H. pylori* infection comprising administering to a subject a vaccine composition of claim 115, such that treatment of *H. pylori* infection occurs.

25 119. A method of claim 118, wherein the treatment is a prophylactic treatment.

120. A method of claim 118, wherein the treatment is a therapeutic treatment.

30 121. A vaccine composition for prevention or treatment of an *H. pylori* infection comprising an effective amount of an *H. pylori* polypeptide of claim 96.

122. A vaccine composition of claim 121, further comprising a pharmaceutically acceptable carrier.

35 123. A vaccine composition of claim 122, wherein the pharmaceutically acceptable carrier is an adjuvant.

- 1129 -

124. A method of treating a subject for *H. pylori* infection comprising administering to a subject a vaccine composition of claim 121, such that treatment of *H. pylori* infection occurs.

5

125. A method of claim 124, wherein the treatment is a prophylactic treatment.

126. A method of claim 124, wherein the treatment is a therapeutic treatment.

10

127. A method for detecting the presence of a *Helicobacter* nucleic acid in a sample comprising:

15

(a) contacting a sample with a nucleic acid of claim 21 under conditions in which a hybrid can form between the probe and a *Helicobacter* nucleic acid in the sample; and

(b) detecting the hybrid formed in step (a), wherein detection of a hybrid indicates the presence of a *Helicobacter* nucleic acid in the sample.

20

128. A method for detecting the presence of a *Helicobacter* nucleic acid in a sample comprising:

(a) contacting a sample with a nucleic acid of claim 37 under conditions in which a hybrid can form between the probe and a *Helicobacter* nucleic acid in the sample; and

25

(b) detecting the hybrid formed in step (a), wherein detection of a hybrid indicates the presence of a *Helicobacter* nucleic acid in the sample.

129. A method for detecting the presence of a *Helicobacter* nucleic acid in a sample comprising: .

30

(a) contacting a sample with a nucleic acid of claim 42 under conditions in which a hybrid can form between the probe and a *Helicobacter* nucleic acid in the sample; and

(b) detecting the hybrid formed in step (a), wherein detection of a hybrid indicates the presence of a *Helicobacter* nucleic acid in the sample.

- 1130 -

130. A method for detecting the presence of a *Helicobacter* nucleic acid in a sample comprising:

- (a) contacting a sample with a nucleic acid of claim 52 under conditions in which a hybrid can form between the probe and a *Helicobacter* nucleic acid in the sample; and
- (b) detecting the hybrid formed in step (a), wherein detection of a hybrid indicates the presence of a *Helicobacter* nucleic acid in the sample.

1/10

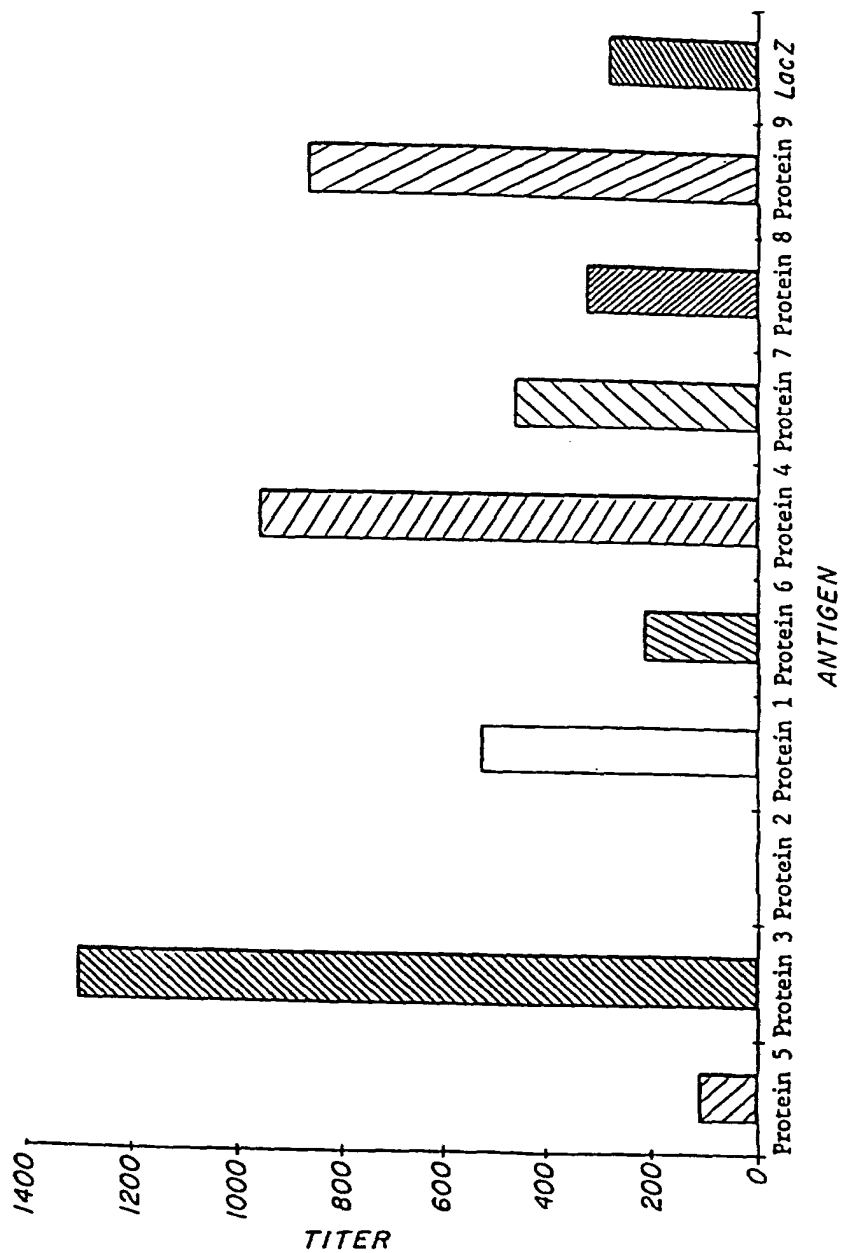


FIGURE 1



2/10

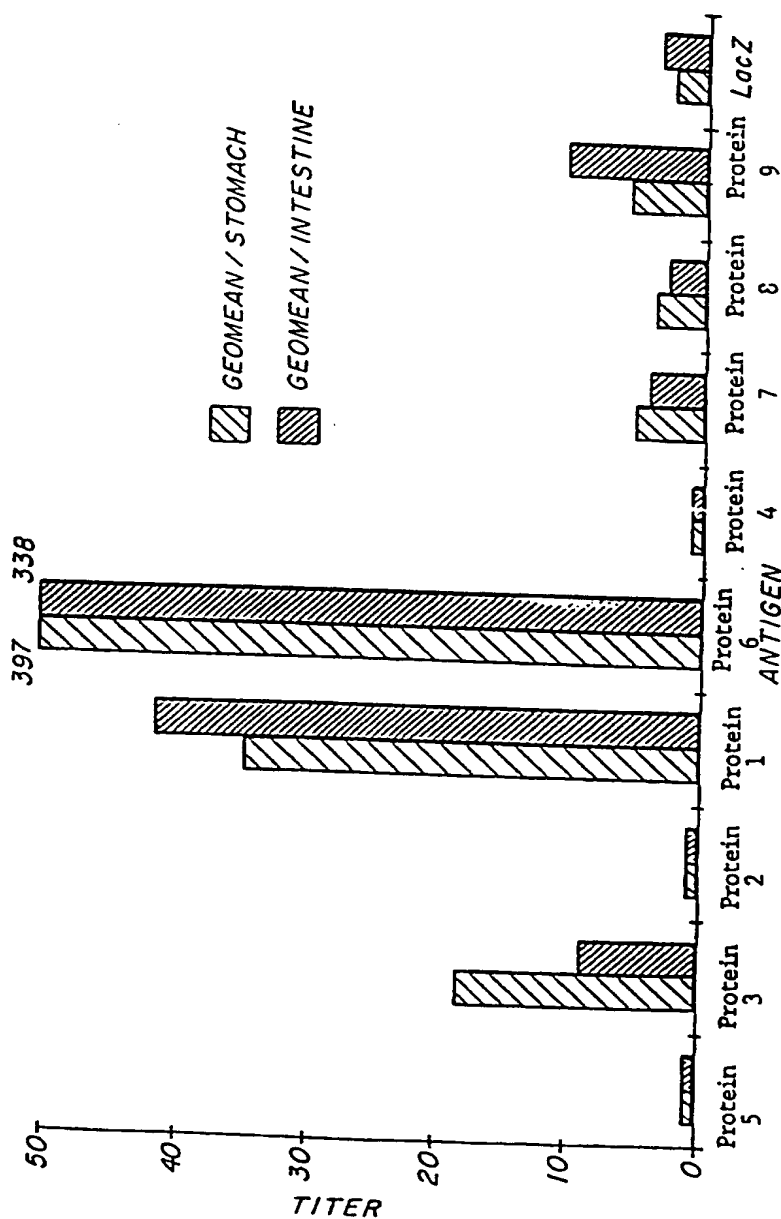


FIGURE 2



3/10

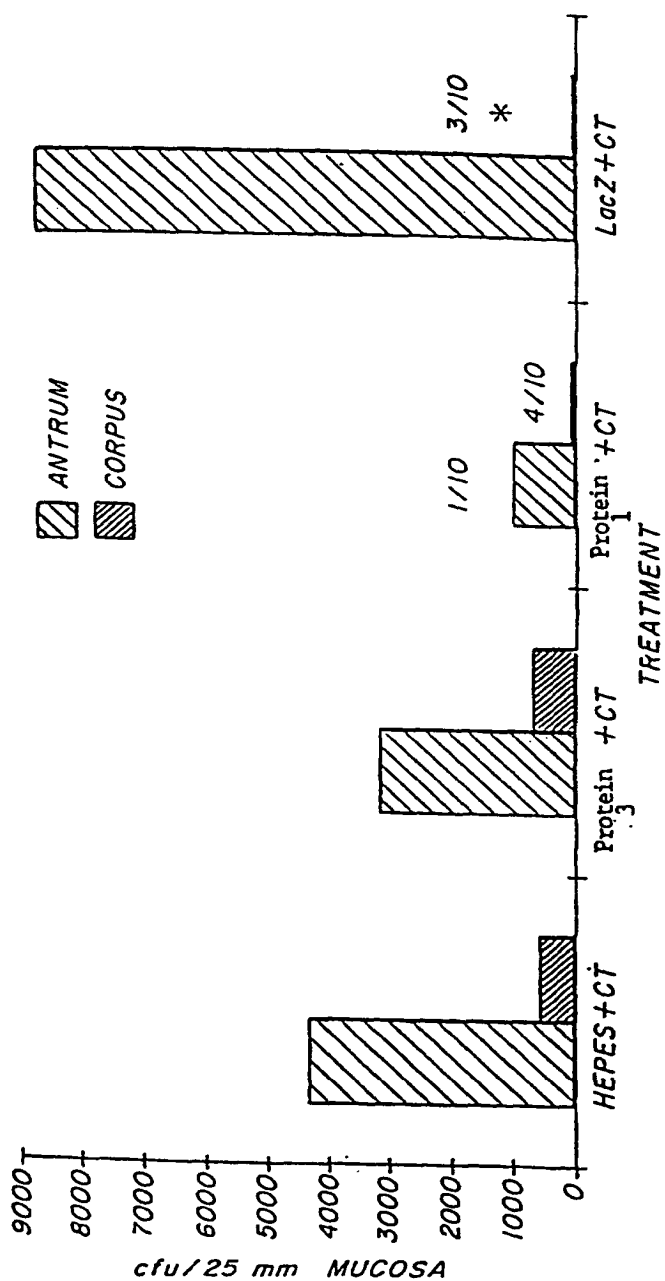


FIGURE 3



4/10

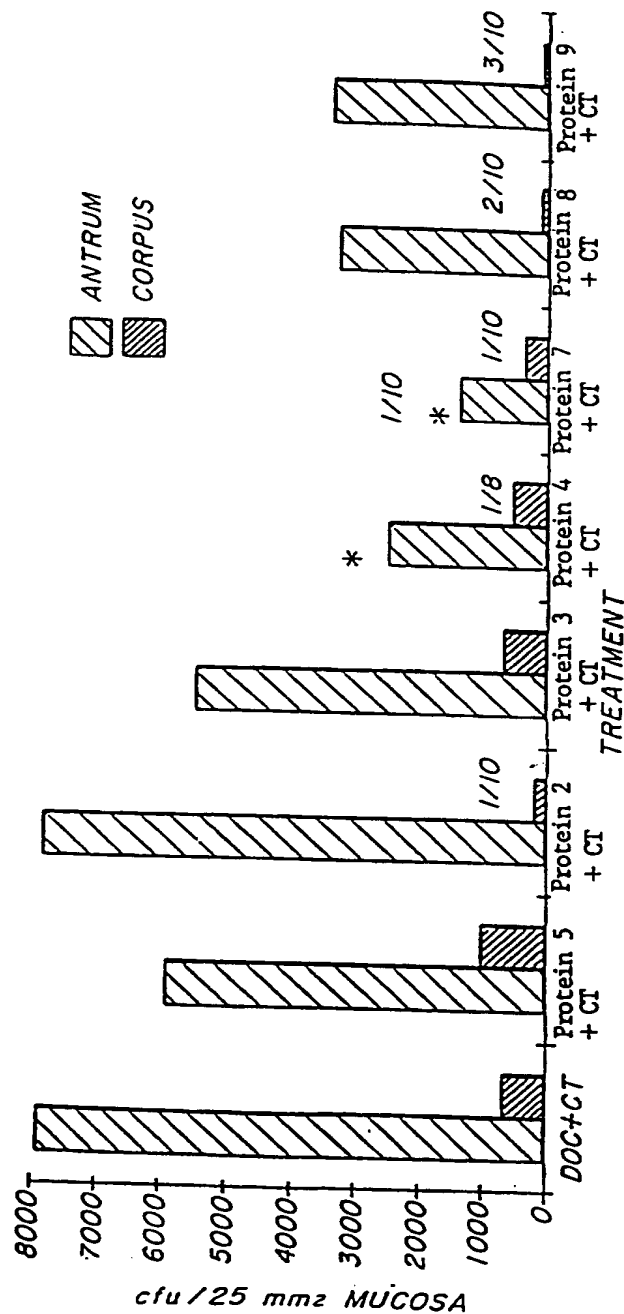
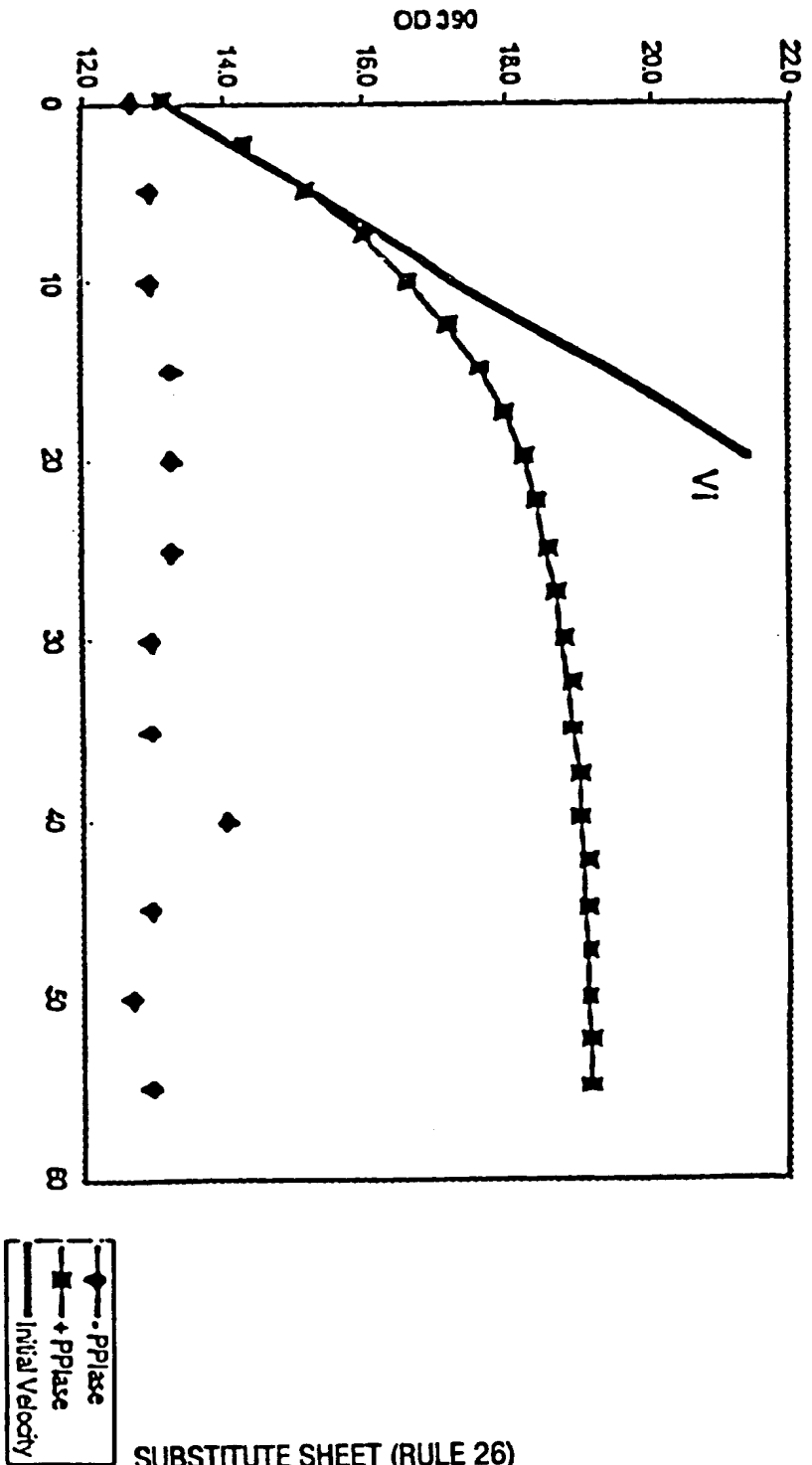


FIGURE 4



5/10

FIGURE 5

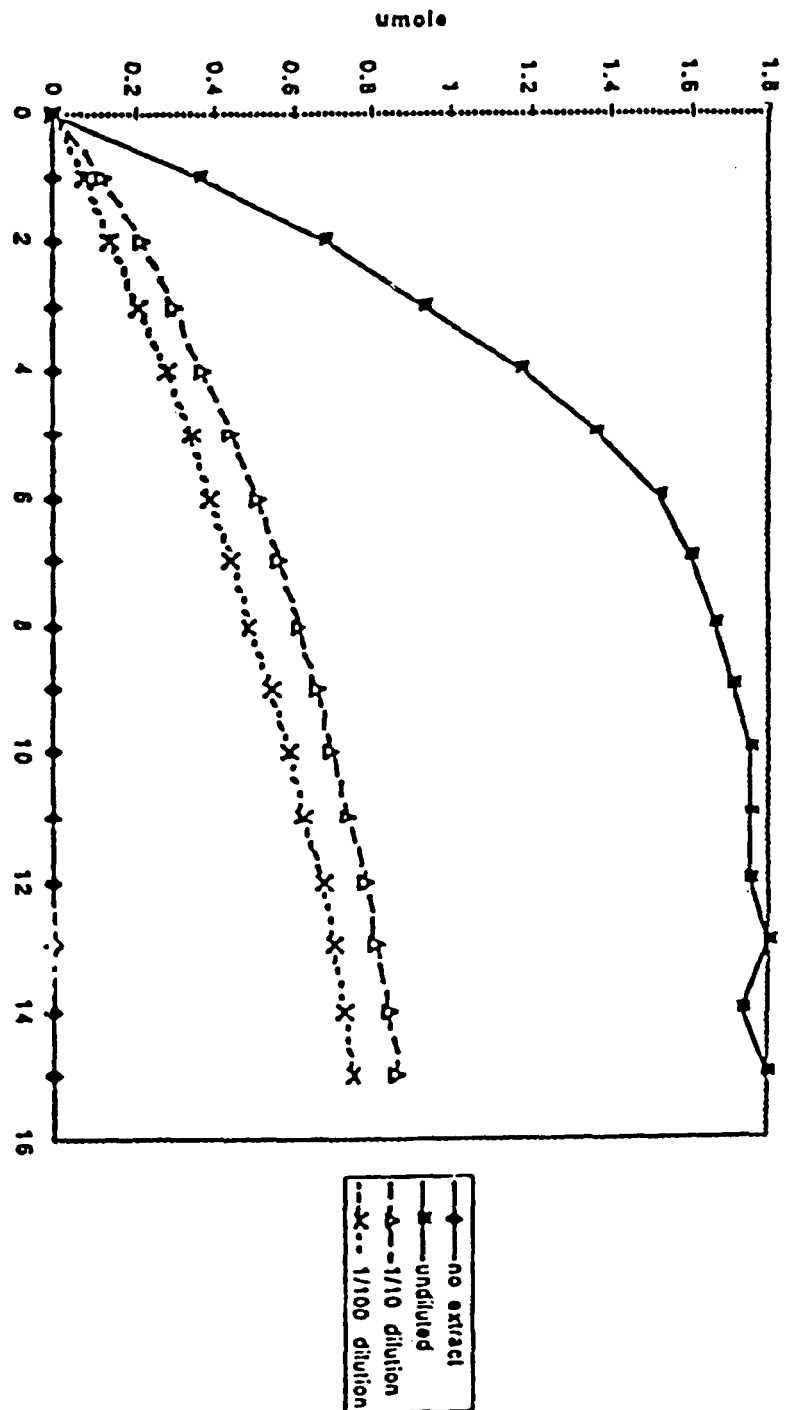


SUBSTITUTE SHEET (RULE 26)



6/10

FIGURE 6



SUBSTITUTE SHEET (RULE 26)



7/10

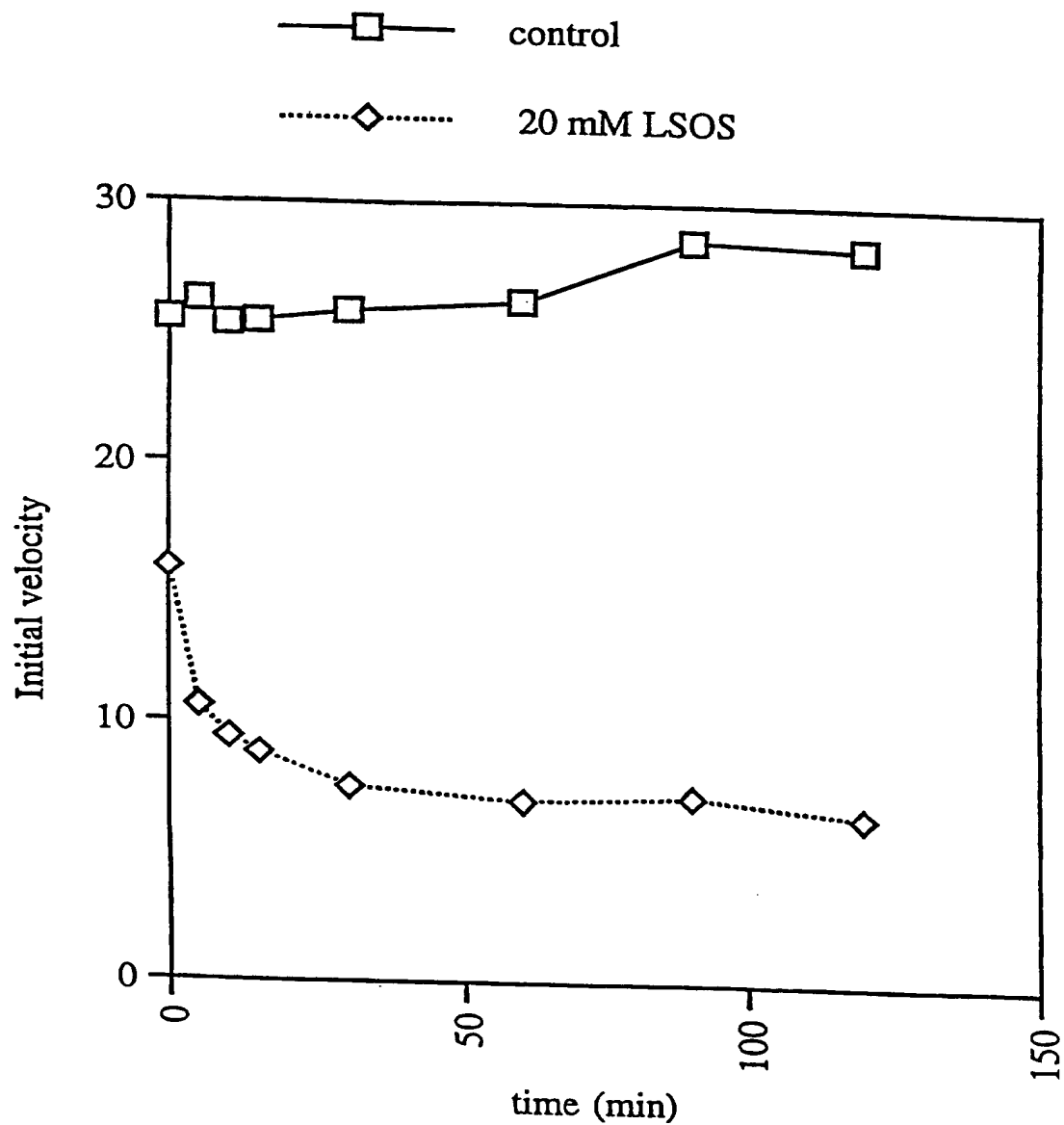


FIGURE 7



aaSeqID#:	BLOCK A	BLOCK B
929	MNGFGAKM	YGVGTDFLYN
938	MNGFGVKM	YGAGTDFLYN
810	LANGFYTKV	YGAGTDVLYN
940	SNGLGVGI	YGAGTDVLYN
945	LNGLGVQV	YGAGTDVLYN
956	MNGIGVQV	YGGGSDLLVN
433	MYGVGLSV	YGAGTDVLYN
777	MYGVDAWA	YGLGIDVLYN
916	INGFGFVV	YGAGTDVLYN
816	SNGLGVQV	YGAGTDVLYN
890	FNELS..L	YGAGTDVLYN
959	PAEVSILI	YGAGTDVLYN
	GYKQFFTH...K KNIGLRYGYFLDY	GYKQFFTH...K KNIGLRYGYFLDY
	GYKQFFGK...K RMFGLRYGYFYDF	GYKQFFGK...K RMFGLRYGYFYDF
	GYKQFFGK...K RNIGLRYGYFFSY	GYKQFFGK...K RNIGLRYGYFFSY
	GYKYFFGK...A RKLGLRHYFFFDY	GYKYFFGK...A RKLGLRHYFFFDY
	GYKQFFGE...S KRWGLRYGYFFDY	GYKQFFGE...S KRWGLRYGYFFDY
	GYKQFFGE...K KRWGLRYGYFFDY	GYKQFFGE...K KRWGLRYGYFFDY
	GYKHFFTK...K KNQGFRTYLYFYDY	GYKHFFTK...K KNQGFRTYLYFYDY
	GYKHFFGK...T KRFGRTYGYYSY	GYKHFFGK...T KRFGRTYGYYSY
	GYKHFVFK...A PQFGHRYGYFFDY	GYKHFVFK...A PQFGHRYGYFFDY
	GYK.WVGKHEET KWFGRMGLFYDL	GYK.WVGKHEET KWFGRMGLFYDL
	GYKYFLGK...K RIIGFRHSLFFGY	GYKYFLGK...K RIIGFRHSLFFGY
	GYKQFFPK...K EHYGFRFYVSLDY	GYKQFFPK...K EHYGFRFYVSLDY
	GYANF...GD.....	GYANF...GD.....
	GYAQF...GT.....	GYAQF...GT.....
	NGASV...GF.....	NGASV...GF.....
	GFSEI...GL.....	GFSEI...GL.....
	NH.....GYIK.SSF.	NH.....GYIK.SSF.
	NH.....FNS..SSDI..WT	NH.....FNS..SSDI..WT
	GYTNF...FNS..ASDV..FT	GYTNF...FNS..ASDV..FT
	NHANL...DG..LCKK NNHL	NHANL...DG..LCKK NNHL
	ASSYY...IMDG..ASQV NNFT	ASSYY...IMDG..ASQV NNFT
	SASLY...KYYTYNDYG MRDARKGSQS	SASLY...KYYTYNDYG MRDARKGSQS
	QLGGV...G.....	QLGGV...G.....
	AYGFFLK NKALODSLR ASSQIPKSYR	AYGFFLK NKALODSLR ASSQIPKSYR
	AYGFFLK NKALODSLR EKLQRKETFI	AYGFFLK NKALODSLR EKLQRKETFI
	NAIF	NAIF
	YVYERS.	YVYERS.
	VFTRK.	VFTRK.
	IFSR.	IFSR.
	LFRR.	LFRR.
	FINDSI	FINDSI
	FINDKT	FINDKT
	FIDN..	FIDN..
	FYES..	FYES..
	....PA	....PA
	AY....	AY....
	WTNDR	WTNDR
	.....	.....

aaSeqID#:	929	.RRR.....	ERTTI.....
	938	.RGT.....	EALDI.....
	810	.YQN.....	RSVDM.....
	940	.YNT.....	KALNF.....
	945	TRKN.....	NKLSV.....
	956	T.KN.....	SKISF.....
	433	..AQ.....	KHSSV.....
	777	..KE.....	GYNTA.....
	916	IFNR.....	ENLHL.....
	816	..NG.....	DKFFA.....
	890	ASQEVERRV	KGLSIFYKDM
	959	.....	TGRTLDADTL
		.....	KRASRHIRK
		.....	.....RA

FIGURE 8



9/10

aaSeqID#:	BLOCK C	BLOCK D	BLOCK E	BLOCK F
929	..	GLFFGQIAGQGW	STNVT NLLSQ.....RPD VKSS	FQFLFDLGV
938	..	GFFAGIQLAGTW	KTNFL DQVDGN.....HLK PKDTS	FQFLFDLGR
810	..	GFFSGIQLAGET	QSTLR DDPNVKL.....HGK INNTH	FQFLDFGMR
940	..	GLFAGVQLGGATW	LSSLR QQIIDNW.....GNAND IHSTN	FQVALNFGVR
945	..	GLFGIQLAGTGW	LNSQY M.....NLTA NNP.....YSAK VNASN	FQFLFNLGLR
956	..	GVFGGIALAGTSW	LNSQY V.....NLATF NNF.....YSAK MNVAN	FQFLFNLGLR
433	..	GFYVGFALAGSSW	VGSGL GMVQSMDFI NNYLTDYRAK MHTSF	FQIPLNFGVR
777	..	GLFVGFGGLGDSF	IVQGE SYLKSQMCI NN.TAGCSAS MNTSY	FQMPVEFGFR
916	..	GFFLGVAIGTSM	GPTNY .....YF KDLAEYRGS FHPSN	FQVLVNGGIR
816	..	GFNLGIAFAG...	...VY DRLSDALLYQ TLLQNTFGK VNUNG	FQFLVDLGV
890	SS	GLVIGMELGASTW	FAS.....NMLTPFNQV KSRTI	FQLOQKFGVR
959	FG	TLILGVNFVGETW	F....YETKIF KQWAKDSLNT YRPNM	FQVMLNVGYR
aaSeqID#:				
929	TNPAK TNFNKH.RLD QG.....I	EFGVKIP	VIA.HKYFATQGSS AS.....	YMRNFSFYVGYSVGF
938	TNFSK IAHQKRSFS QG.....I	EGLKIPV	LY.HTYYQSEGV AK.....	YRRAFSFGVGNIGF
810	MNFGK LDGKSNRHQ HT.....V	EFGVVVP	TIY.NTYYKSAGTT VK.....	YERPYSVMWSYGSE
940	TNFAE PKRFAKKFHN QGVISQSV	EFGIKVP	LIN.QAYLNSAGAD VS.....	YRRLTYFYINYIMGF
945	TNLAT AKKKSERSA QH.....GV	ELGKIP	TIN.TNYYSLGTQ LE.....	YRRLYSVYLYNVFAY
956	MNLAK NKKASDHAA QH.....GV	ELGVKIP	TIN.TNYYSLGTQ LQ.....	YRRLYSVYLYNVFAY
433	VNVDR .....HN.....GF	EMGLKIP	LAV.NSFYETHGK LNA.....SLF	FKRLVMFNVSYYVSF
777	SNFSK .....HS.....GI	EVGFKLP	LFT.NQFYKERGV DSV.....DVF	YKRNFSIYFNMYMINL
916	LG... .....TK HQ.....GF	EIGLKIQ	TIR.NNYYTASADN VPE.....GT TYRFT	FHRPYAFYWRIVVSF
816	LGNEH NQ.....F	GFGIKIP	TYFENHYTSMNNTS NNSENLVKL RFEYGINSL LYQVD	FRNYSVYFNITYVSF
890	..FSS DEYDIDRYCD ENYLGSSV	ELGVKVP	AFKNYYSDDYGDK LD.....	YKRVSVYLYNITYNF
959	YRFSR YKN.....WAI	EFGARIP	FL.INDYFKTP....LYTLHFKR NI	SVYLTSTYDF

FIGURE 8 (CONTINUED)



10/10

aaSeqID# :

810	MQQQLTYLNAGNVFFNAMNKA	LEKNGTA	TANSTSSTSGATGSDG
925	MQFQKTLLSLSLLFLSYCI	AEENGAY	ASVGFEYSISHAVEHN
929	LKNHSFKKTIALSLLASMSLCN	AEEDGAF	FVIDYQTSLARQELKN
938	LQNFVFNKKWLIYSSLLPLFFLNPLM	AEDDGFF	MGVSYQTS LAVQRVDN
956	KKPFYSLSLASSLLN	AEDNGFF	ISAGYQIGEEAQMVKV
945	LH	AEDNGFF	VSAGYQIGEAQMVKV
940	MIKKAKKFIPLFLIGSLL	AEDNGWY	MSVGYQIGGTQQFINN
890	MKKVLLLTLSLSLSFWLH	AERNIFY	LGLNFAEGSYIQGQGS
777	MKKIFLGMALAFVSVM	AEKSGAF	LGGGFQYSNLENQNTT

FIGURE 9



## INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US97/05223

## A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) : C12Q 1/68; A01N 43/04; A61K 31/70

US CL : 435/6; 514/44

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/6; 514/44

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

genbank, dialog

search terms: specific SEQ ID No's and helicobacter pylori

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	BUKANOV, N.O. et al. Ordered cosmid library and high-resolution physical-genetic map of <i>Helicobacter pylori</i> strain NCTC11638. Molecular Microbiology. February 1994, Vol. 11, No. 3, page 509-523, see whole article.	1, 5-7, 14-15, 21-23, 39, 40-41, 43-44, 61-63, 67-69, 77-79, 83-85
A	TAYLOR, D.E. et al. Construction of a <i>Helicobacter pylori</i> genome map and demonstration of diversity at the genome level. Journal of Bacteriology. November 1992, Vol. 174, No. 21, pages 6800-6806, especially pages 6802-6803.	1, 5-7, 14-15, 21-23, 39, 40-41, 43-44, 61-63, 67-69, 71, 77-79, 83-85



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:	"T" later document published after the international filing date or priority date and not in conflict with the application but cited to underscore the principle or theory underlying the invention
"A" document defining the general state of the art which is not considered to be of particular relevance	"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E" earlier document published on or after the international filing date	"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"Z" document member of the same patent family
"O" document referring to an oral disclosure, use, exhibition or other means	
"P" document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search

14 JULY 1997

Date of mailing of the international search report

05 AUG 1997

Name and mailing address of the ISA/US  
Commissioner of Patents and Trademarks  
Box PCT  
Washington, D.C. 20231

Facsimile No. (703) 305-3230

Authorized officer

GINNY PORTNER

Telephone No. (703) 308-0196

Form PCT/ISA/210 (second sheet)(July 1992)\*

# INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US97/05223

## Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:  
because they relate to subject matter not required to be searched by this Authority, namely:
2. ☐ Claims Nos.:  
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
3. ☐ Claims Nos.:  
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

## Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:  
Please See Extra Sheet

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☒ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:  
1-130(in part, elected SEQ ID NOs as noted in response to 206)
4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

☐  
☒

The additional search fees were accompanied by the applicant's protest.  
No protest accompanied the payment of additional search fees.

## INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US97/05223

### BOX II. OBSERVATIONS WHERE UNITY OF INVENTION IS LACKING (Continued):

This International Search Authority has found 309 inventions claimed in the International Application covered by the claims indicated below:

This application contains the following inventions or groups of inventions which are not so linked as to form a single inventive concept under PCT Rule 13.1. In order for all inventions to be searched, the appropriate additional search fees must be paid.

Group I, claims 1-13, 21-79 and 127-130 are directed to no fewer than 620 DNAs, vectors containing the different DNAs, organisms transformed with the DNAs, DNA encoding fragments of the polypeptides encoded by the no fewer than 620 different DNAs, vaccines and methods of producing the polypeptides encoded by the no fewer than 620 different DNAs.

Group II, claims 14-20 and 80-126 are drawn to no fewer than 620 different polypeptides encoded by a subset of the encoding DNAs mentioned in Group I.

This application contains claims directed to more than one species of the generic invention. These species are deemed to lack Unity of Invention because they are not so linked as to form a single inventive concept under PCT Rule 13.1. In order for more than one species to be searched, the appropriate additional search fees must be paid. The species are as follows:

Group I contains a separate DNA species for each sequence mentioned. Therefore, there is a minimum of 620 species. Group II contains at least one polypeptide for each DNA sequence mentioned. Therefore, this is a minimum of 620 species in Group II.

For either Group that applicant elects, a total of 10 (ten) specified sequences will be searched and no more than 4 (four) specified sequences will be searched for each additional fee paid; if no additional fee is paid and no election indicated the first 10 sequences will be searched.

and it considers that the International Application does not comply with the requirements of unity of invention (Rules 13.1, 13.2 and 13.3) for the reasons indicated below:

The polypeptide encoding DNAs, vectors containing them, organisms transformed with them and methods of polypeptide production using them are materially different from and are therefore independent from the polypeptides of Group II. Additionally, none of the products or methods of Group I is needed to make the polypeptides of Group II.

The species listed above do not relate to a single inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, the species lack the same or corresponding special technical features for the following reasons: There is no relationship between or among the various nucleotide and amino acid sequences mentioned in the claims.

